Vol. ……
ISSN : 2580-6181 (Print), 2599-2481 (Online)
Available online at:
http://ejournal.upi.edu/index.php/tarbawy/index

ANALISIS BUTIR PENILAIAN TENGAH SEMESTER PENDIDIKAN AGAMA ISLAM KELAS X DI SMAN 10 YOGYAKARTA 2022/2023

Amalia Nurlitasari*, Maemunah

UIN Sunan Kalijaga Yogyakarta, Indonesia
E-mail: 22204011003@student.uin-suka.ac.id

Abstract. *The knowledge and skills of teachers in conducting question point analysis are still relatively low. Teachers still tend to make questions according to core competencies and basic competencies without considering the quality of the test. This study aims to see the quality of question items and examples of questions that are feasible with improper through the Midterm Assessment (PTS) class X Islamic Religious Education (PAI) subject at SMAN 10 Yogyakarta. This research is a type of quantitative research and uses JASP Program version 16.0. The results showed that reliability in the "High" category. The most p mean value in the "Medium" category of 56%. The most D value in the "Enough" category of 44% and the most distractor in the "Accepted" category of 128 out of 200 questions. Through this calculation, it can be seen the characteristics of decent and unfeasible questions.*
Keywords*: Item Analysis, Islamic Religious Educator (IRE)*

Abstrak. *Pengetahuan dan keterampilan guru dalam melakukan analisis butir soal masih tergolong rendah. Guru masih cenderung membuat soal sesuai dengan kompetensi inti dan kompetensi dasar tanpa mempertimbangkan kualitas tes. Penelitian ini bertujuan melihat kualitas butir soal dan contoh soal yang layak dengan tidak layak melalui Penilaian Tengah Semester (PTS) kelas X mata pelajaran Pendidikan Agama Islam (PAI) di SMAN 10 Yogyakarta. Penelitian ini merupakan jenis penelitian kuantitatif dan menggunakan Program JASP versi 16.0. Adapun hasil penelitian menunjukan bahwa Reliabilitas pada kategori "Tinggi". Nilai mean p paling banyak pada kategori "Sedang" sebesar 56%. Nilai D paling banyak pada kategori "Cukup" sebesar 44% dan distractor paling banyak pada kategori "Diterima" sebesar 128 dari 200 soal. Melalui perhitungan tersebut maka dapat dilihat karakteristik soal yang layak dan tidak layak.*

Kata Kunci*: Analisis Butir, Pendidikam Agama Islam*

PENDAHULUAN

Penilaian hasil belajar yang dilakukan oleh guru telah dijelaskan dalam Permendikbud RI Nomor 104 Tahun 2014 tentang Penilaian Hasil Belajar Oleh Pendidik dan Satuan Pendidikan pada Pendidikan Dasar dan Pendidikan Menengah Pasal 1, penilaian hasil belajar oleh pendidik adalah pengumpulan informasi tentang capaian pembelajaran peserta didik dalam kometensi sikap spiritual dan sosial, pengetahuan, dan keterampilan yang diperoleh dari pembelajaran (Permendikbud No 104, 2014). Dalam mengukur capaian pembelajaran saat ini, kurikulum merdeka belajar mendesain asesmen atau penilaian menjadi tiga kategori, yaitu asesmen diagnostik, asesmen formatif, dan asesmen sumatif (Maulinda, 2022). Penilaian hasil belajar yang sampai sekarang masih sering atau dominan dilakukan ialah pada aspek penilaian sumatif (Anisah, 2022). Penilaian sumatif digunakan sebagai acuan melihat keberhasilan Kegiatan Belajar Mengajar (KBM) peserta didik baik setelah kurang lebih 3 bulan pembelajaran atau disebut dengan Penilaian Tengah Semester (PTS) maupun setelah satu semester pembelajaran disebut dengan Penilaian Akhir Tahun (PAT) (Barokah, 2019).

Melihat pentingnya penilaian sumatif dalam mengukur capaian pembelajaran peserta didik, SMAN 10 Yogyakarta yang merupakan salah satu sekolah negeri di Yogyakarta menggunakan tes sumatif sebagai proses penilaian hasil belajar peserta didik. Penilaian yang dilakukan salah satunya yaitu Penilaian Tengah Semester (PTS) berbentuk Pilihan Ganda yang dilakukan setelah 3 bulan proses belajar mengajar dilaksanakan. Peneliti menekankan analisis butir soal Penilaian Tengah Semester (PTS) sebab saat penelitian dilakukan SMA N 10 Yogyakarta akan melakukan PTS yang berbentuk pilihan ganda sejumlah 40 soal. Peserta didik diwajibkan mengerjakan soal ujian PTS yang dibuat dan disusun oleh tim guru pengampu mata pelajaran. Oleh karenanya soal ujian yang dibuat guru harus memiliki kualitas mutu yang baik dan mampu memberikan hasil yang tepat sebagai salah satu sarana dalam melakukan proses asesmen agar dapat mengetahui kemampuan peserta didik dengan baik (Himawan & Nurgiyantoro, 2022).

Idealnya penilaian yang disusun oleh guru mampu memberikan banyak informasi mengenai kemampuan siswa (Setiawan et al., 2020). Akan tetapi, persoalan muncul saat guru menyusun butir soal masih banyak ditemukan kelemahan dalam proses penyusunannya sehingga menghasilkan tes yang tidak memiliki daya ukur yang valid. Instrumen tes yang tidak memiliki daya ukur yang valid, tidak akan memberikan informasi apapun mengenai kemampuan peserta didik (Aripiani, 2014). Analisis instrumen butir soal sangat penting untuk dilakukan oleh guru karena terkadang butir soal dibuat terlalu sulit atau terlalu mudah sehingga menyebabkan guru sulit untuk membedakan karakteristik peserta didik dari segi kemampuan dan keberhasilan peserta didik (Nur & Palobo, 2018).

Selain itu pengetahuan dan keterampilan guru dalam melakukan analisis butir soal masih tergolong rendah (Prihatin, 2022). Beberapa hasil penelitian terdahulu mengenai analisis butir soal menunjukan guru hanya membuat soal sesuai dengan kompetensi inti dan kompetensi dasar saja tanpa melakukan analisis kualitas tiap item butir soal dengan rincian sebagai berikut: *Pertama,* dari hasil wawancara penelitian ditemukan bahwa guru belum mampu mengembangkan kemampuannya dalam penyusunan tes yang baik. Guru hanya mengambil dari buku pelajaran, bank soal, dan internet lalu setelah selesai dibuat, guru langsung mengujikan kepada peserta didik (Rokhyati, 2011). *Kedua,* guru jarang

melakukan analisis butir soal sebab kurang fahamnya guru dalam mengalisis tiap butir soal. Guru melakukan analisis butir soal menggunakan cara manual melalui excel (Mia, 2014). *Ketiga,* informasi yang didapat melalui wawancara ialah soal PTS Genap kelas V diujikan oleh guru pengampu mata pelajaran tanpa melakukan analisis baik secara kualitatif maupun kuantitatif (Ardianto, 2019).

Hal ini juga di konfirmasi oleh Wakil Kepala Bidang Kurikulum PAI di SMA N 10 Yogyakarta berdasarkan wawancara pada tanggal 3 Mei 2023 bahwa dalam pembuatan soal guru telah melaksanakan telaah soal sebelum dibagikan kepada peserta didik berupa pengecekan bentuk soal dan kesesuaian soal dengan kisi-kisi maupun materi yang diajarkan. Namun, guru belum melakukan analisis butir soal yang digunakan dalam pembuatan soal PAT maupun PTS. Ketidakmampuan dan kurangnya wawasan mengenai analisis butir soal ini menyebabkan guru tidak bisa memperoleh timbal balik berupa informasi untuk membedakan karakteristik peserta didik dari segi kemampuan dan keberhasilan peserta didik. Dengan demikian sudah dapat dipastikan bahwa keberhasilan siswa dalam kegiatan belajar mengajar tentunya tidak dapat diukur secara tepat.

Agar penilaian dapat diukur secara tepat maka perlu adanya analisis butir soal. Analisis butir soal diartikan sebagai langkah dalam mengkaji item pertanyaan tiap butir soal yang akan diujikan agar diperoleh soal ujian yang berkualitas (Sudjana, 2016). Dalam menganalisis butir soal terdapat dua pendekatan dalam menganalisis data hasil uji coba soal, yaitu dengan menggunakan teori klasik dan menggunakan teori respon butir (Nur & Palobo, 2018). Teori klasik umumnya menggunakan teknik kualitatif melalui *expert judgement* untuk melihat kesesuaian antara butir soal dengan materi pelajaran dalam kurikulum yang selanjutnya dituangkan pada kisi-kisi soal. Sedangkan teori respon butir menggunakan teknik kuantitatif yang berisi reliabilitas soal, indeks kesukaran, indeks daya beda, dan keefektifan distraktor (Rokhyati, 2011).

Memilih PAI sebagai mata pelajaran yang diteliti tentu bukan tanpa sebab. Pembelajaran Agama Islam disekolah sudah ada sejak seseorang mulai bersekolah baik di bangku Taman Kanak-kanak hingga perkuliahan untuk mengajarkan berbagai aspek seperti Alqur'an Hadits, Aqidah, Akhlak, Fiqh, Dan Tarikh (Taufik & Halimah, 2019). Maka dari itu, amat penting untuk dilakukan analisis butir soal di tingkat menengah atas yang sudah dapat membedakan mana yang baik dan buruk agar guru mendapatkan informasi mengenai sejauh mana kemampuan keagamaan peserta didik.

Melalui teknik kuantitatif guru akan mendapatkan umpan balik sehingga dapat digunakan sebagai acuan dalam pengambilan keputusan soal yang akan diujikan. Hasil analisis butir soal dengan teknik kuantitatif berkaitan dengan reliabilitas, indeks kesukaran (ITK), indeks daya beda (IDB) dan keefektifan distraktor. Hasil reliabilitas soal berfungsi untuk mengetahui apakah soal PTS kelas X dapat digunakan secara konsisten dalam waktu yang berbeda (Hanifah, 2014). Reliabilitas dilihat berdasarkan skor *Alpha Cronbach* dengan klasifikasi 0,00-0,20 (sangat rendah); 0,21-0,40 (rendah); 0,41-0,70 (sedang); 0,71-0,90 (tinggi); 0,91-1.00 (sangat tinggi) (Arikunto, 2016). Bilangan yang menunjukan sukar atau mudahnya suatu soal disebut dengan indeks kesukaran (ITK). Indeks kesukaran diperoleh dari skor perhitungan program JASP dengan klasifikasi 0,00-0,30 (sukar); 0,31-0,70 (sedang); dan 0,71-1,00 (mudah) (Arikunto, 2016). Hasil dari indeks tingkat

kesukaran soal digunakan untuk melihat mutu item butir soal PTS (Sudjana, 2016).

Indeks daya beda (IDB) digunakan untuk mengetahui perbedaan kemampuan peserta didik yang telah mampu menguasai dengan yang belum menguasai materi. dapat dinyatakan memenuhi syarat layak apabila indeks daya beda diatas 0,20 (Hanifah, 2014). Keefektifan distraktor/ pengecoh merupakan pilihan distraktor pada setiap soal pilihan ganda. Pada Penilaian Tengah Semester (PTS) kelas X mata pelajaran PAI di SMA N 10 Yogyakarta menggunakan 4 distraktor dan satu kunci. Dalam memenuhi kriteria keefektifan distraktor, maka perlu 5% peserta yang terkecoh (Subali, 2014). Setelah melakukan beberapa proses analisis butir soal, maka soal yang masuk kategori layak dapat langsung digunakan, soal dengan kategori kurang layak dapat direvisi, sedangkan soal dengan kategori tidak layak dapat digugurkan sehingga harus dihapus atau diganti dengan soal lainnya (Mania et al., 2020).

Berdasarkan bukti penilaian terdahulu yang mana guru lebih condong untuk melakukan analisis soal melalui teknik kualitatif, maka pada peneliti mengusung kebaharuan penelitian dengan lebih menekankan pada teknik kuantitatif mata pelajaran Pendidikan Agama Islam (PAI). Maka dari itu, penelitian ini memiliki kontribusi penelitian bidang Pendidikan Agama Islam (PAI) yang membahas analisa butir soal Penilaian Tengah Semester sehingga dapat diketahui kualitas soal yang digunakan untuk mendapat informasi mengenai karakteristik peserta didik. menggunakan program JASP dan hasil kajian penelitian ini dapat dijadikan referensi bagi penelitian selanjutnya.

METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif yang dilakukan melalui analisis empiris dengan metode deskriptif kuantitatif. Menurut Hadad Nawawi dalam (Rotama et al., 2020), analisis deskriptif kuantitatif merupakan cara dalam menggambarkan suatu keadaan penelitian. Deskriptif kuantitatif digunakan untuk mendeskripsikan hasil perhitungan analisis butir soal pada program JASP versi 16.0 dibantu dengan Microsoft Excel.

Teknik pengumpulan data menggunakan wawancara dengan subjek informan Wakil Kepala Bidang Kurikulum Ibu Mar'atul Allamah, S.Ag. dan berkas dokumen soal Penilaian Tengah Semester (PTS) Kelas X Mata Pelajaran Pendidikan Agama Islam (PAI) yang berjumlah 40 soal yang dibuat oleh guru pengampu mata pelajaran. Namun pada saat perhitungan analisis, soal nomor 20 gugur karena nilai menunjukan negatif sehingga tidak memenuhi syarat.

Penelitian ini berlokasi di SMA N 10 Yogyakarta tepatnya di Jl. Gadean No.5, Ngupasan, Kec. Gondomanan, Kota Yogyakarta.

Populasi dalam penelitian ini berupa soal PTS mata pelajaran PAI kelas X. Sampel yang digunakan berupa jawaban peserta didik kelas X IPS 1 yang berjumlah sebanyak 36 peserta didik. Peneliti memilih kelas X IPS 1 sebab kelas tersebut merupakan kelas tinggi atau unggulan.

HASIL PENELITIAN DAN PEMBAHASAN

Analisis Reliabilitas

Reliabilitas soal merupakan hal yang sangat penting dalam pengukuran. Keajegan hasil tes yang didapat oleh orang yang sama ketika diuji ulang meskipun dengan kondisi waktu yang berbeda menjadi acuan kepercayaan tes (Arikunto, 2016). Menurut Zainal Arifin, suatu soal dinilai memiliki reliabilitas tinggi apabila memberikan konsistensi. Semakin reliabel sebuah soal ujian maka

hasil ujian tersebut akan mempunyai hasil yang sama saat dipakai kembali (Zainal Arifin, 2016).

Reliabilitas dilihat berdasarkan skor *Alpha Cronbach* dengan klasifikasi 0,00-0,20 (sangat rendah); 0,20-0,40 (rendah); 0,40-0,70 (sedang); 0,70-0,90 (tinggi); 0,90-1.00 (sangat tinggi) (Arikunto, 2016). Berikut hasil perhitungan reliabilitas soal yang diperoleh melalui program JASP:

Gambar 1. Hasil Reliabilitas

Frequentist Scale Reliability Statistics

Estimate	Cronbach's α	Guttman's λ2	Average interitem correlation
Point estimate	0.864	0.880	0.149
95% CI lower bound	0.787	0.733	0.069
95% CI upper bound	0.918	0.942	0.269

Note. Variables V9 and V10 correlated perfectly.

Berdasarkan gambar 1. Hasil reliabilitas melalui program JASP peneliti menggunakan dua rumus dalam melihat reliabilitas soal yaitu Cronbach's dengan nilai 0,864 dan Guttman dengan nilai 0,880. Keduanya menunjukan hasil yang perbedaannya tidak terlalu jauh dan sama-sama diatas 0,85. Artinya kestabilan hasil soal ujian menurut klasifikasi reliabilitas dapat dikatakan berada pada kriteria **tinggi** sebab nilai Alpha sebesar **0,864.** Sehingga kapanpun soal ujian tersebut digunakan, akan memeberikan hasil yang relatif sama.

Butir soal ujian yang baik akan memiliki tingkat reliabilitas yang tinggi, menghasilkan hasil yang ajeg, maka dari itu apabila ditemukan soal yang tidak reliabel maka dapat langsung diganti dengan soal baru.

Analisis Indeks Tingkat Kesukaran

Melalui Indeks Tingkat Kesukaran (ITK), guru mampu mengetahui seberapa mudah atau sukarnya soal yang diujikan kepada peserta didik (Burhan, 2020). Indeks kesukaran pada hasil perhitungan JASP dilihat melalui nilai *Mean*. Dalam penelitian ini untuk mengetahui indeks kesukaran selain menggunakan JASP juga bisa melalui Microsoft Excel dengan formula "If". Hasil yang muncul pun menunjukan hasil yang sama.

Indeks kesukaran yang digunakan dalam penelitian ini menurut Arikunto terbagi menjadi tiga klasifikasi sebagai berikut: 0,00-0,30 (sukar); 0,31-0,70 (sedang); dan 0,71-1,00 (mudah). Jika butir soal memiliki angka kesukaran 0,00 maka soal tersebut sangat sukar karena tidak ada peserta didik yang menjawab dengan benar. Begitu pula sebaliknya, jika suatu soal memiliki angka kesukaran 1,00 maka soal tersebut tergolong mudah karena semua peserta ujian mampu menjawab dengan benar. Soal yang baik ialah soal yang tidak terlalu mudah atau tidak terlalu sukar. Soal yang terlalu mudah, tidak merangsang peserta didik untuk memecahkannya. Sebaliknya, jika butir soal terlalu sukar, akan menyebabkan peserta didik menjadi putus asa dan tidak memiliki semangat untuk mencoba mengerjakannya kembali (Arikunto, 2016).

Berikut hasil perhitungan analisis Indeks Tingkat Kesukaran menggunakan JASP:

V33	0.857	0.874	0.496	0.639
V34	0.868	0.884	0.054	0.333
V35	0.857	0.874	0.511	0.694
V36	0.858	0.875	0.557	0.917
V37	0.868	0.883	0.064	0.694
V38	0.869	0.884	0.020	0.639
V39	0.861	0.878	0.372	0.917
V40	0.861	0.878	0.403	0.944

Note. The following items were reverse scaled: V1, V22, V28, V37.

Gambar 2. Hasil Indeks Kesukaran

Berdasarkan gambar 2. Tingkat kesulitan soal Penilaian Tengah Semester kelas X Mata Pelajaran PAI dapat dilihat pada tabel *Mean P* kemudian di interpretasikan melalui klasifikasi tingkat kesukaran soal. Dilihat dari klasifikasi tingkat kesukaran soal, yang masuk dalam kategori soal **"Sukar"** berada pada nomor 3, 22, 25 atau sebesar **8%**. Kategori soal **"Sedang"** pada nomor 1, 2, 5, 8, 16, 17, 18, 19, 21, 23 ,24, 26, 29, 30, 31, 32, 33, 34, 35, 37 dan 38 atau sebesar **56%**. Kategori **"Mudah"** berada pada nomor 4, 6, 9, 10, 11, 12, 13, 14, 15, 27, 28, 36, 39, 40 atau sebesar **36%.**

Jika soal sukar artinya tingkat kemampuan peserta didik rendah, dengan kata lain jika soal diujikan pada peserta dengan kemampuan rendah, maka tingkat kesukaran soal akan menjadi tinggi. Sebaliknya jika soal mudah artinya tingkat kemampuan peserta didik tinggi, dengan kata lain tingkat kesukaran soal akan menjadi rendah ketika soal diujikan pada kelompok berkemampuan tinggi, kedua tingkat kesukaran soal didefinisikan sebagai proporsi peserta didik dalam grup yang menjawab soal benar (Nurcahyo, 2017).

Soal yang memiliki indeks tingkat kesukaran yang imbang maka bisa dikatakan soal tersebut baik (Zainal Arifin, 2016). Hal ini dijelaskan lebih lanjut oleh Sudjana yang mengatakan bahwa, "Perbandingan antara soal yang mudah-sedang-sukar dapat dibuat 3-4-3, yang memiliki arti, 30% soal berkategori mudah, 40% soal berkategori sedang, dan 30% soal berkategori sukar". Berdasarkan penjelasan tersebut, dapat diketahui bahwa soal-soal yang mudah atau sukar, bukan berarti tidak boleh digunakan. Hendaknya soal tersebut ditindaklanjuti oleh pembuat soal. Butir soal yang sukar diteliti ulang sehingga dapat diketahui mengapa soal tersebut masuk dalam kategori mudah maupun sukar. (Sudjana, 2016).

Hasil analisis butir soal PTS kelas X di SMA N 10 Yogyakarta menunjukan belum mencapai 3-4-3. Kategori sukar mendapat jatah paling sedikit atau tidak memenuhi 30% dibandingkan dengan kategori mudah yang sudah diatas 30%. Sedangkan untuk kategori sedang mendapatkan jatah hingga diatas 50%.

Butir soal yang sukar dijawab oleh peserta didik dapat disebabkan oleh beberapa hal. Pengecekan terhadap kunci jawaban perlu dilakukan. Apabila kunci jawaban sudah benar, sulitnya soal dimungkinkan karena materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya, sehingga kompetensi minimum yang harus dikuasai peserta didik belum tercapai. Faktor lain yang berpengaruh adalah materi yang diukur tidak cocok

Frequentist Individual Item Reliability Statistics

Item	If item dropped		Item-rest correlation	mean
	Cronbach's α	Guttman's λ2		
V1	0.868	0.884	0.010	0.222
V2	0.863	0.880	0.250	0.361
V3	0.866	0.882	0.076	0.167
V4	0.854	0.872	0.703	0.861
V5	0.856	0.874	0.550	0.722
V6	0.862	0.879	0.305	0.972
V7	0.858	0.875	0.557	0.917
V8	0.855	0.873	0.573	0.750
V9	0.859	0.877	0.435	0.889
V10	0.859	0.877	0.435	0.889
V11	0.859	0.877	0.420	0.889
V12	0.859	0.877	0.402	0.833
V13	0.859	0.876	0.506	0.917
V14	0.859	0.876	0.450	0.889
V15	0.857	0.874	0.529	0.833
V16	0.861	0.878	0.327	0.778
V17	0.860	0.877	0.379	0.611
V18	0.856	0.874	0.518	0.722
V19	0.863	0.880	0.243	0.806
V21	0.855	0.872	0.570	0.583
V22	0.865	0.881	0.120	0.861
V23	0.855	0.872	0.594	0.722
V24	0.853	0.871	0.639	0.639
V25	0.866	0.882	0.043	0.111
V26	0.863	0.880	0.278	0.639
V27	0.861	0.878	0.331	0.889
V28	0.865	0.881	0.115	0.111
V29	0.861	0.878	0.342	0.556
V30	0.862	0.879	0.278	0.806
V31	0.864	0.880	0.213	0.750
V32	0.860	0.876	0.386	0.639

ditanyakan dengan menggunakan bentuk soal yang diberikan dan pernyataan atau

kalimat soal terlalu panjang (Rokhyati, 2011).

Analisis Indeks Daya Beda

Menurut Sudaryono dalam fitrianawati, daya beda butir tes adalah kemampuan suatu butir soal dapat membedakan peserta tes yang telah menguasai materi yang ditanyakan dan peserta tes yang belum menguasai materi yang ditanyakan (Fitrianawati, 2017). Indeks daya beda mampu membedakan mana peserta didik yang pandai (kelompok atas) dan yang kurang pandai (kelompok bawah) dari skor total (Zainal Arifin, 2016). Peserta didik yang pandai akan lebih banyak menjawab benar, sedangkan peserta didik yang kurang pandai akan menjawab tidak benar. Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminasi atau D (Anas Sudijono, 2016).

Dalam program JASP daya beda dapat dilihat melalui *Item-rest correlation*. Pada penelitian ini peneliti menggunakan indeks daya pembeda menurut Arikunto yang mengklasifikasikan daya pembeda soal ke dalam lima kategori yaitu 0,00-0,20 (Jelek); 0,21-0,40 (Cukup); 0,41-0,70 (Baik); diatas 0,70 (Baik Sekali). Angka indeks diskriminasi berkisar antara Indeks daya beda yang baik berkisar antara 0,40-0,70 pada kategori (Baik). (Arikunto, 2016).

Selain menggunakan JASP untuk mengetahui daya beda, pada penelitian ini juga menggunakan bantuan Microsoft Excel dengan formula "If". Melalui Microsoft Excel, hasil analisis menunjukan lebih mudah untuk melihat manakah peserta didik yang termasuk dalam kelompok atas maupun kelompok bawah. Beberapa pakar lebih banyak menggunakan presentase sebesar 27% dari masing-masing kelompok (Anas Sudijono, 2016). Hasil yang muncul pun menunjukan hasil yang sama. Hasil perhitungan IDB menggunakan program JASP sebagai berikut:

Gambar 3. Hasil Indeks Daya Beda

Frequentist Individual Item Reliability Statistics

	If item dropped			
Item	Cronbach's α	Guttman's λ2	Item-rest correlation	mean
V1	0.868	0.884	0.010	0.222
V2	0.863	0.880	0.250	0.361
V3	0.866	0.882	0.076	0.167
V4	0.854	0.872	0.703	0.861
V5	0.856	0.874	0.550	0.722
V6	0.862	0.879	0.305	0.972
V7	0.858	0.875	0.557	0.917
V8	0.855	0.873	0.573	0.750
V9	0.859	0.877	0.435	0.889
V10	0.859	0.877	0.435	0.889
V11	0.859	0.877	0.420	0.889
V12	0.859	0.877	0.402	0.833
V13	0.859	0.876	0.506	0.917
V14	0.859	0.876	0.450	0.889
V15	0.857	0.874	0.529	0.833
V16	0.861	0.878	0.327	0.778
V17	0.860	0.877	0.379	0.611
V18	0.856	0.874	0.518	0.722
V19	0.863	0.880	0.243	0.806
V21	0.855	0.872	0.570	0.583
V22	0.865	0.881	0.120	0.861
V23	0.855	0.872	0.594	0.722
V24	0.853	0.871	0.639	0.639
V25	0.866	0.882	0.043	0.111
V26	0.863	0.880	0.278	0.639
V27	0.861	0.878	0.331	0.889
V28	0.865	0.881	0.115	0.111
V29	0.861	0.878	0.342	0.556
V30	0.862	0.879	0.278	0.806
V31	0.864	0.880	0.213	0.750
V32	0.860	0.876	0.386	0.639
V33	0.857	0.874	0.496	0.639
V34	0.868	0.884	0.054	0.333
V35	0.857	0.874	0.511	0.694
V36	0.858	0.875	0.557	0.917
V37	0.868	0.883	0.064	0.694
V38	0.869	0.884	0.020	0.639
V39	0.861	0.878	0.372	0.917
V40	0.861	0.878	0.403	0.944

Note. The following items were reverse scaled: V1, V22, V28, V37.

Berdasarkan gambar 3. Hasil daya beda Penilaian Tengah Semester kelas X Mata Pelajaran PAI dapat dilihat pada tabel *Item-rest correlation* kemudian di interpretasikan melalui klasifikasi Indeks Daya Beda. Dilihat dari klasifikasi daya beda, yang masuk dalam kategori soal **"Jelek"** berada pada nomor 3, 6, 9, 10, 22, 25, 28, 37, 38, 40 atau sebesar **26%**. Kategori soal **"Cukup"** pada nomor 1, 7, 11, 12, 13, 14, 16, 17, 18, 19, 26, 27, 30, 31, 34, 36, 39 atau sebesar **44%**. Kategori **"Baik"** berada pada nomor 2, 4, 5, 15, 29, 32, 35 atau sebesar **18%**. Kategori **"Baik Sekali"** berada pada nomor 8, 21, 23, 24, 33 atau sebesar **13%**. Soal terbanyak berada pada kategori cukup. Namun,

menjadi perhatian bagi guru mata pelajaran sebagai pembuat soal untuk melihat kembali karakteristik soal pada kategori jelek sebab presentase soal jelek cukup tinggi yaitu 26%.

Analisis Keefektifan Distraktor

Soal yang baik harus memiliki pengecoh yang relative homogen sehingga tidak mudah ditebak oleh peserta didik (Ratnaningsih et al., 2013).

Pengecoh berfungsi dengan baik apabila memiliki daya tarik yang besar bagi peserta didik yang kurang menguasai dan memahami materi. Justru Ketika pengecoh tidak dipilih sama sekali oleh peserta didik, artinya pengecoh tersebut jelek, terlalu menyesatkan (Sudaryono, 2011). Sehingga dalam menganalisis Langkah pertama yang harus dilakukan adalah menghitung banyaknya jumlah peserta didik yang memilih masing-masing pilihan jawaban. Kemudian gunakan formula = (Jumlah peserta didik yang memilih jawaban A, B, C, D, atau E/ total jumlah peserta didik yang mengerjakan) untuk menghitung tingkat keefektifan distraktor/ pengecoh seperti pada gambar 4.

Gambar 4. Hasil keefektifan distraktor

NO	A	B	C	D	E
1	0,0%	19,4%	0,0%	72,2%	2,8%
2	2,8%	11,1%	16,7%	36,1%	22,2%
3	44,4%	8,3%	16,7%	16,7%	13,9%
4	86,1%	8,3%	0,0%	2,8%	2,8%
5	11,1%	11,1%	5,6%	63,9%	0,0%
6	0,0%	0,0%	91,7%	2,8%	0,0%
7	91,7%	2,8%	5,6%	0,0%	0,0%
8	8,3%	72,2%	5,6%	2,8%	5,6%
9	5,6%	0,0%	83,3%	2,8%	2,8%
10	88,9%	0,0%	5,6%	2,8%	2,8%
11	88,9%	5,6%	2,8%	0,0%	0,0%
12	5,6%	8,3%	80,6%	0,0%	2,8%
13	8,3%	88,9%	0,0%	0,0%	0,0%
14	88,9%	11,1%	0,0%	0,0%	0,0%
15	2,8%	0,0%	13,9%	75,0%	0,0%
16	77,8%	5,6%	11,1%	0,0%	5,6%
17	2,8%	16,7%	11,1%	55,6%	5,6%
18	2,8%	69,4%	22,2%	0,0%	2,8%
19	11,1%	0,0%	5,6%	2,8%	69,4%
20	13,9%	25,0%	27,8%	22,2%	5,6%
21	19,4%	2,8%	11,1%	52,8%	8,3%
22	30,6%	19,4%	11,1%	13,9%	22,2%
23	11,1%	5,6%	66,7%	11,1%	0,0%
24	16,7%	61,1%	8,3%	2,8%	8,3%
25	11,1%	5,6%	5,6%	55,6%	13,9%
26	2,8%	8,3%	58,3%	11,1%	13,9%
27	0,0%	0,0%	11,1%	80,6%	0,0%
28	88,9%	11,1%	0,0%	0,0%	0,0%
29	19,4%	52,8%	22,2%	2,8%	0,0%
30	0,0%	5,6%	11,1%	2,8%	72,2%
31	13,9%	2,8%	5,6%	2,8%	66,7%
32	8,3%	2,8%	58,3%	22,2%	2,8%
33	8,3%	0,0%	19,4%	55,6%	5,6%
34	33,3%	44,4%	8,3%	5,6%	2,8%
35	5,6%	13,9%	63,9%	11,1%	0,0%
36	5,6%	88,9%	0,0%	0,0%	2,8%
37	8,3%	19,4%	30,6%	19,4%	19,4%
38	11,1%	11,1%	13,9%	55,6%	0,0%
39	91,7%	2,8%	0,0%	5,6%	0,0%
40	0,0%	0,0%	5,6%	86,1%	0,0%

Berdasarkan gambar 4. Kolom yang berwarna kuning merupakan kunci jawaban. Keefektifan pengecoh sangat

dibutuhkan dalam soal dimana peserta didik salah dalam memilih jawaban. Hendaknya pengecoh disesuaiakan dengan materi pembelajaran(Fitrianawati, 2017). Langkah terakhir dalam melakukan analisis keefektifan distraktor ialah dengan menggunakan formula "if" sesuai dengan indeks yang telah ditentukan yaitu D>5% maka diterima, 5%>D>0% maka direvisi, dan D=0% maka ditolak. Berikut hasil analisis keefektifan distraktor menggunakan Microsoft Excel:

Gambar 5. Keterangan keefektifan distraktor

NO	A	B	C	D	E
1	Ditolak	Diterima	Ditolak	Diterima	Revisi
2	Revisi	Diterima	Diterima	Diterima	Diterima
3	Diterima	Diterima	Diterima	Diterima	Diterima
4	Diterima	Diterima	Ditolak	Revisi	Revisi
5	Diterima	Diterima	Diterima	Diterima	Ditolak
6	Ditolak	Ditolak	Diterima	Revisi	Ditolak
7	Diterima	Revisi	Diterima	Ditolak	Ditolak
8	Diterima	Diterima	Diterima	Revisi	Diterima
9	Diterima	Ditolak	Diterima	Revisi	Revisi
10	Diterima	Ditolak	Diterima	Revisi	Revisi
11	Diterima	Diterima	Revisi	Ditolak	Ditolak
12	Diterima	Diterima	Diterima	Ditolak	Revisi
13	Diterima	Diterima	Ditolak	Ditolak	Ditolak
14	Diterima	Diterima	Ditolak	Ditolak	Ditolak
15	Revisi	Ditolak	Diterima	Diterima	Ditolak
16	Diterima	Diterima	Diterima	Ditolak	Diterima
17	Revisi	Diterima	Diterima	Diterima	Diterima
18	Revisi	Diterima	Diterima	Ditolak	Revisi
19	Diterima	Ditolak	Diterima	Revisi	Diterima
20	Diterima	Diterima	Diterima	Diterima	Diterima
21	Diterima	Revisi	Diterima	Diterima	Diterima
22	Diterima	Diterima	Diterima	Diterima	Diterima
23	Diterima	Diterima	Diterima	Diterima	Ditolak
24	Diterima	Diterima	Diterima	Revisi	Diterima
25	Diterima	Diterima	Diterima	Diterima	Diterima
26	Revisi	Diterima	Diterima	Diterima	Diterima
27	Ditolak	Ditolak	Diterima	Diterima	Ditolak
28	Diterima	Diterima	Ditolak	Ditolak	Ditolak
29	Diterima	Diterima	Diterima	Revisi	Ditolak
30	Ditolak	Diterima	Diterima	Revisi	Diterima
31	Diterima	Revisi	Diterima	Revisi	Diterima
32	Diterima	Revisi	Diterima	Diterima	Revisi
33	Diterima	Ditolak	Diterima	Diterima	Diterima
34	Diterima	Diterima	Diterima	Diterima	Revisi
35	Diterima	Diterima	Diterima	Diterima	Ditolak
36	Diterima	Diterima	Ditolak	Ditolak	Revisi
37	Diterima	Diterima	Diterima	Diterima	Diterima
38	Diterima	Diterima	Diterima	Diterima	Ditolak
39	Diterima	Revisi	Ditolak	Diterima	Ditolak
40	Ditolak	Ditolak	Diterima	Diterima	Ditolak

Berdasarkan gambar 5. Pengecoh yang ditolak artinya pilihan jawaban tersebut tidak ada yang memilih sehingga tidak dapat berfungsi secara efektif sebagai pengecoh. Pengecoh yang direvisi artinya berada pada taraf kurang dari 5% namun diatas 0% maka hendaknya guru mengganti atau menelaah kembali pilihan jawaban tersebut. Lalu apabila pengecoh tersebut diterima, artinya pilihan jawaban pengecoh tersebut berfungsi dengan baik.

Berdasarkan penjelasan tersebut, dapat disimpulkan bahwa pengecoh ialah pilihan jawaban yang bukan merupakan kunci jawaban. Meskipun pilihan jawaban tersebut ialah opsi yang salah, namun pilihan jawaban salah tersebut harus bisa berperan sesuai keefektifannya, baik sebagai perusak, penjebak, maupun pengecoh terhadap beberapa peserta ujian (Himawan & Nurgiyantoro, 2022). Analisis butir soal secara kuantitatif juga dapat dilakukan menggunakan program komputer. Namun karena keterbatasan program JASP dalam penganalisisan keefektifan distraktor, maka untuk mengetahui keefektifannya maka peneliti menggunakan Microsoft Excel.

Contoh soal PTS kelas X di SMA N 10 Yogyakarta dengan kategori layak dan tidak layak.

22.	Pernyataan dibawah ini yang **bukan** merupakan contoh dari maukuf alaih atau nazir adalah…

	A. Ibu ani menyerahkan sebidang tanah kepada bapak ali B. H. Hasan merupakan ketua panitia pembangunan masjid al ikhlash C. Iuran jariah dari jamaah untuk sarana majlis ta'lim diterima hj. Aisyah D. Sumbangan wakaf untuk klinik kesehatan diterima oleh h. Amir E. Panitia pembangunan madrasah tsanawiyah al-mahbubiyah

Berdasarkan hasil analisis soal, soal no. 22 dalam Indeks Tingkat Kesukaran menunjukan angka 0,14 yang masuk dalam kategori **"sukar".** indeks daya beda sebesar 0,12 yang artinya Jelek. Mayoritas kelompok atas yang menjawab benar hanya 1 dan kelompok bawah yang menjawab benar hanya 2 peserta didik menyebabkan soal tersebut tidak dapat membedakan peserta didik yang memahami materi dengan yang tidak sehingga masuk kedalam kategori **"tidak layak".** Namun hal ini berarti soal tersebut memiliki kefektifan distraktor yang bagus pada pilihan jawaban. Keefektifan distraktor pada no. 22 berada pada kategori **"Diterima"** yang menunjukan setiap pilihan jawaban telah dipilih oleh peserta didik diatas 5%.

Jika dilihat dari soal, maka kemungkinan yang menyebababkan soal tersebut tidak layak kelima pilihan jawaban tersebut menggunakan pilihan kata yang ambigu, tidak menggunakan kata yang mudah dipahami sehingga yang ada hanya keefektifan distraktor saja namun aspek ITK dan IDB tidak terpenuhi.

21.	Adanya wakif, nazir, mauquf bih, dan sighat merupakan…
	A. Hukum wakaf B. Sunnah wakaf C. Wajib wakaf D. Rukun wakaf E. Bentuk wakaf

Soal nomor 21 dikatakan berada pada kategori layak sebab hasil analisis ITK sebesar 0,72 masuk dalam kategori **"sedang"**. Artinya soal tersebut tidak terlalu sulit dan tidak terlalu mudah. Indeks daya beda sebesar 0,80 masuk kedalam kategori **"baik sekali"** yang artinya soal tersebut mampu membedakan peserta didik yang memahami materi dengan yang tidak. Namun kaitannya dengan keefektifan distraktor, pilihan jawaban A, C, D, E masuk dalam kategori **"diterima"** sedangkan opsi B menunjukan kategori **"revisi"**. Pilihan jawaban yang direvisi bukan berarti harus dibuang, revisi memiliki makna boleh digunakan namun guru harus menelaah kembali pilihan kata opsi B yang sesuai dengan lingkup pada opsi pilihan A, C, D, dan E.

KESIMPULAN
(Century 12 Bold)

Berdasarkan analisis butir soal dari segi Reliabilitas, Indeks Tingkat Kesukaran (ITK), Indeks Daya Beda (IDB), dan Keefektifan Distraktor didapat hasil sebagai berikut:

1. Dilihat dari klasifikasi reliabilitas dapat dikatakan berada pada kriteria **"Tinggi"** sebab nilai Alpha sebesar 0,864 sehingga kapanpun soal ujian tersebut digunakan, akan memeberikan hasil yang relatif sama.
2. Dilihat dari klasifikasi tingkat kesukaran soal, yang masuk dalam kategori soal **"Sukar"** berada pada nomor 3, 22, 25 atau sebesar **8%**. Kategori soal **"Sedang"** pada nomor 1, 2, 5, 8, 16, 17, 18, 19, 21, 23 ,24, 26, 29, 30, 31, 32, 33, 34, 35, 37 dan 38 atau sebesar **56%**. Kategori **"Mudah"** berada pada nomor 4, 6, 9, 10, 11, 12, 13,

14, 15, 27, 28, 36, 39, 40 atau sebesar **36%.**

3. Dilihat dari klasifikasi daya beda, yang masuk dalam kategori soal **"Jelek"** berada pada nomor 3, 6, 9, 10, 22, 25, 28, 37, 38, 40 atau sebesar **26%**. Kategori soal **"Cukup"** pada nomor 1, 7, 11, 12, 13, 14, 16, 17, 18, 19, 26, 27, 30, 31, 34, 36, 39 atau sebesar **44%**. Kategori **"Baik"** berada pada nomor 2, 4, 5, 15, 29, 32, 35 atau sebesar **18%**. Kategori **"Baik Sekali"** berada pada nomor 8, 21, 23, 24, 33 atau sebesar **13%**.
4. Keefektifan distraktor menunjukan total sebanyak **42 dari 200** pilihan jawaban masuk dalam kategori **"ditolak"** dan sisanya pada kategori diterima dan direvisi. Hal ini menjadi perhatian guru untuk mengganti soal yang ditolak agar pilihan kata yang ditolak pada suatu nomor dapat disesuaikan dengan konteks yang sesuai. Soal yang sudah masuk kategori **"direvisi"** sebanyak **30** dari **200** dapat ditelaah kembali dan yang sudah masuk kategori **"diterima"** sebanyak **128** dari **200** dapat dipertahankan.

REFERENSI
(Century 12 Bold)

Anas Sudijono. (2016). Anas Sudijono, Pengantar Evaluasi Pendidikan. *RajaGrafindo Persada*, *Cet. Ke-16*.

Anisah, G. (2022). KERANGKA KONSEP ASSESSMENT OF LEARNING, ASSESSMENT FOR LEARNING, DAN ASSESSMENT AS LEARNING SERTA PENERAPANNYA PADA PEMBELAJARAN. *AL-AUFA: JURNAL PENDIDIKAN DAN KAJIAN KEISLAMAN*, *3*(2). https://doi.org/10.32665/alaufa.v3i2.1201

Ardianto, N. (2019). *ANALISIS BUTIR SOAL PENILAIAN TENGAH SEMESTER GENAP KELAS V TEMA 6 PANAS DAN PERPINDAHANNYA SD DABIN III KECAMATAN TEGAL BARAT KOTA TEGAL*. Universitas Negeri Semarang.

Arikunto, S. (2016). *Dasar-dasar evaluasi pendidikan* (Revisi, Cet. Ke-8). Bumi Aksara.

Aripiani, T. D. (2014). *ANALISIS ALAT PENILAIAN PEMBELAJARAN UJIAN AKHIR SEMESTER MATA PELAJARAN PRODUKTIF KEAHLIAN KELAS X TATA BOGA DI SMK SHANDY PUTRA BANDUNG*. Universitas Pendidikan Indonesia.

Barokah, M. (2019). Manajemen penilaian sumatif pada ranah kognitif pembelajaran PAI kelas X semester ganjil di SMA Negri 2 Pontianak tahun pelajaran 2017/ 2018. *Al-Idarah: Jurnal Kependidikan Islam*, *9*(2).

Burhan, N. (2020). Penilaian Pembelajaran Bahasa Berbasis Kompetensi. *Didaktika: Jurnal Kependidikan*, *9*(1).

Fitrianawati, M. (2017). Peran Analisis Butir Soal Guna Meningkatkan Kualitas Butir Soal, Kompetensi Guru Dan Hasil Belajar Peserta Didik. *JPT : Jurnal Pendidikan Tematik*, *2*(3).

Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *SOSIO E-KONS*, *6*(1).

Himawan, R., & Nurgiyantoro, B. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul menggunakan program ITEMAN. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, *8*(1).

Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). ANALISIS BUTIR SOAL UJIAN AKHIR SEKOLAH. *Al Asma : Journal of Islamic Education*, *2*(2). https://doi.org/10.24252/asma.v2i2.16569

Maulinda, U. (2022). Pengembangan Modul Ajar Berbasis Kurikulum Merdeka. *Tarbawi*, *5*(2).

Mia, A. I. (2014). *ANALISIS BUTIR SOAL DALAM BUKU AJAR ILMU PENGETAHUAN ALAM KELAS V DI MADRASAH IBTIDAIYAH AL MAARIF 02 PALANG SUKOREJO PASURUAN*. UIN MAULANA MALIK IBRAHIM.

Nur, A. S., & Palobo, M. (2018). Pelatihan Analisis Butir Soal Berbasis Komputerisasi Pada Guru SD. *MATAPPA: Jurnal Pengabdian Kepada Masyarakat*, *1*(1). https://doi.org/10.31100/matappa.v1i1.79

Nurcahyo, F. A. (2017). Aplikasi IRT dalam Analisis Aitem Tes Kognitif. *Buletin Psikologi*, *24*(2). https://doi.org/10.22146/buletinpsikologi.25218

Permendikbud No 104. (2014). Salinan Lampiran Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 104 Tahun 2014 Tentang Penilaian Hasil Belajar oleh Pendidik Pada Pendidikan Dasar dan Pendidikan Menengah. *Pedoman Evaluasi Kurikulum*.

Prihatin, R. P. (2022). *Analisis Soal SKI Kelas X MAN 2 Bantul Yogyakarta Pada Penilaian Akhir Tahun 2020/2021*. UIN SUNAN KALIJAGA .

Ratnaningsih, D. J., Isfarudi, I., & Soleiman, N. (2013). Analisis Butir Soal Pilihan Ganda Ujian Akhir Semester Mahasiswa Di Universitas Terbuka Dengan Pendekatan Teori Tes Klasik. *Jurnal Pendidikan Terbuka Dan Jarak Jauh*, *12*(2).

Rokhyati. (2011). *KARAKTERISTIK SECARA KUALITATIF DAN KUANTITATIF SOAL ULANGAN AKHIR SEMESTER GENAP BAHASA INDONESIA KELAS XII SMA NEGERI DI KABUPATEN PURBALINGGA TAHUN PELAJARAN 2010/2011*. UNIVERSITAS NEGERI YOGYAKARTA.

Rotama, A. D., Budiutomo, T. W., & Bowo, A. N. A. (2020). ANALISIS BUTIR SOAL PENILAIAN TENGAH SEMESTER MATA PELAJARAN PPKn KELAS VII DI SMP MUHAMMADIYAH 7 YOGYAKARTA. *Academy of*

Education Journal, *11*(01). https://doi.org/10.47200/aoej.v11i01.314

Setiawan, M. A., Susongko, P., & Hayati, M. N. (2020). Pendeteksian DIF pada Perangkat Tes Objektif Penilaian Akhir Semester IPA dengan Mengunakan Permodelan Rasch. *PSEJ (Pancasakti Science Education Journal)*, *5*(2).

Subali, B. (2014). Analisis Soal Baik Kualitatif Maupun Kuantitatif. *Workshop Item Development Dosen Poltekes Kebidanan Politeknik Kesehatan Surakarta*, *5*.

Sudaryono. (2011). Implementasi Teori Responsi Butir (Item Response Theory) pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan Dan Kebudayaan*, *17*(6).

Sudjana, N. (2016). Penilaian Hasil Proses Belajar Mengajar. In *Penilaian dan Hasil Belajar Mengajar*.

Taufik, A., & Halimah, I. (2019). Buku Guru Pendidikan Agama Islam dan Budi Pekerti Kelas X SMA/SMK. In *Direktorat Jenderal Pendidikan Agama Islam Kementerian Agama Indonesia*.

Zainal Arifin. (2016). Evaluasi Pembelajaran. *Jurnal Pendidikan*, *03*.