

KARAKTERISTIK BUTIR SOAL UJIAN AKHIR SEMESTER MATA KULIAH DASAR-DASAR PENYULUHAN

Nurul Huda
email: nurul@ecampus.ut.ac.id

Abstrak

Dalam menyusun soal tes untuk mengukur capaian hasil belajar, perlu diketahui karakteristik butir soal karena kualitas butir soal akan tercermin dari karakteristik butir soalnya. Tulisan ini bertujuan untuk menentukan: (1) sejauhmana karakteristik butir soal UAS pada mata kuliah Dasar-dasar penyuluhan pertanian ditinjau dari segi validitas, reliabilitas, dan tingkat kesukaran; serta (2) sejauhmana tingkat kualitas soal UAS pada mata kuliah Dasar-dasar Penyuluhan. Yang menjadi subjek dalam tulisan ini adalah mata kuliah Dasar-dasar Penyuluhan Pertanian pada program studi agribisnis UT. Data dikumpulkan dari hasil respon mahasiswa pada ujian akhir semester (UAS) masa ujian 2023.2, yang bersumber dari aplikasi pusat pengujian. Analisis data dilakukan dengan analisis butir soal tes yang meliputi validitas, reliabilitas, dan tingkat kesukaran dengan pendekatan kuantitatif. Dari hasil analisis, beberapa temuan yang dihasilkan adalah: (1) Tingkat validitas butir soal mata kuliah Dasar-dasar Penyuluhan Pertanian tergolong rendah; sedangkan tingkat reliabilitas tergolong sedang. Tingkat kesukaran butir soal masuk ke dalam kategori sedang = 88%, sukar = 4% mudah = 5,3%, sangat sukar = 0%, dan sangat mudah = 2,7%, (2) secara umum kualitas butir soal mata kuliah Dasar-dasar Penyuluhan tergolong kurang baik. Saran yang dapat disampaikan yaitu; (1) butir soal tes yang perlu direvisi diharapkan segera direvisi untuk memperbaiki kualitas butir soal yang bersangkutan; (2) butir soal yang ditolak sebaiknya tidak diikutkan dalam soal tes selanjutnya. Jika memungkinkan karena secara umum hasil analisis butir soal mata kuliah ini kurang baik, maka disarankan untuk mengganti set soal dengan yang baru.

Kata kunci: karakteristik, butir soal, ujian akhir semester, dasar-dasar penyuluhan

PENDAHULUAN

Hasil belajar merupakan elemen penting dalam proses pembelajaran karena dapat mengetahui capaian hasil belajar mahasiswa setelah menyelesaikan proses pembelajaran. Pengukuran hasil belajar ini merupakan kegiatan evaluasi yang dilakukan untuk menentukan apakah program pembelajaran yang telah dilaksanakan berhasil atau tidak.

Salah satu pengukuran hasil belajar adalah melalui tes atau ujian berbentuk soal objektif dengan pilihan berganda. Parameter karakteristik butir soal pilihan ganda yang baik yaitu valid dan reliabel. Validitas berdasarkan isi tes (validitas isi) sangat penting untuk memvalidasi skor tes tersebut, sedangkan reliabilitas tes pilihan ganda diestimasi dengan menggunakan formula Kuder Richardson-20 (KR-20) karena skor jawaban tes tersebut bersifat dikotomi. Formula KR-20 digunakan untuk mengestimasi indeks reliabilitas data kategorikal dan dikotomus.

Selain validitas dan reliabilitas, karakteristik butir soal tes juga dilihat dari tingkat kesukaran, daya beda, dan keberfungsian pengecoh. Tingkat kesukaran merujuk pada besarnya persentase peserta tes yang menjawab dengan benar. Mardapi (2017) menyatakan bahwa tingkat kesukaran merupakan proporsi peserta yang menjawab dengan benar dimana besarnya indeks tingkat kesukaran berkisar antara 0,00 sampai dengan 1,00. Semakin kecil indeks tingkat kesukarannya, maka butir soal tersebut semakin sulit. Sebaliknya, semakin besar indeks tingkat kesukarannya, maka butir soal tersebut akan semakin mudah. Amalia & Widayati (2012) menyatakan bahwa soal yang baik adalah soal yang tidak terlalu mudah atau tidak terlalu sukar. Menurut Azwar (2016), formula untuk menghitung indeks tingkat kesukaran butir soal adalah $P = Np/N$, dimana P = indeks kesukaran butir soal, Np = jumlah peserta menjawab benar, serta N = jumlah seluruh peserta tes. Kriteria tingkat kesukaran butir soal yang baik adalah apabila memiliki indeks kesukaran butir soal antara 0,30 sampai dengan 0,70. Apabila indeks kesukaran butirnya antara 0,10 sampai dengan 0,29 atau antara 0,71 sampai dengan 0,90, maka butir soal tersebut perlu direvisi. Selanjutnya apabila indeks kesukaran butirnya kurang dari 0,10 atau lebih dari 0,90, maka butir soal tersebut ditolak/didrop (Mardapi, 2017; Azwar 2016; dan Kartowagiran, 2012). Seiring dengan perkembangan teknologi, analisis soal kini dapat dilakukan secara lebih mudah dan efisien menggunakan perangkat lunak seperti Anates (Sari et al., 2020). Software ini membantu guru menganalisis kualitas butir soal berdasarkan berbagai parameter, seperti validitas, reliabilitas, tingkat kesulitan, daya pembeda, dan kualitas pengecoh (Kustati & Amelia, 2024)

Daya beda menunjukkan kemampuan suatu butir soal tes membedakan peserta tes yang berkemampuan tinggi dengan peserta tes yang berkemampuan rendah. Menurut Solichin (2017) daya beda soal adalah kemampuan suatu soal untuk membedakan peserta yang pandai dengan yang kurang pandai. Indeks daya beda butir soal dihitung dengan rumus $d = (NiT / NT) - (NiR / NR)$, dimana: d = indeks daya beda, NiT = banyaknya penjawab benar dari kelompok tinggi, NT = banyaknya penjawab dari kelompok tinggi, NiR = banyaknya peserta yang menjawab benar dari kelompok rendah, serta NR = banyaknya peserta dari kelompok rendah (Azwar, 2016). Indeks daya beda berkisar antara 0,00 sampai dengan 1,00. Indeks daya beda yang mendekati 0 menunjukkan bahwa butir tersebut memiliki daya beda yang rendah. Mardapi (2017) dan Kartowagiran (2012) menyatakan bahwa indeks daya beda dikatakan baik jika lebih besar dari 0,30 sehingga butir soal tersebut layak diterima. Butir soal yang memiliki indeks daya beda sebesar 0,10 sampai dengan 0,29 perlu direvisi, sementara butir soal yang memiliki indeks daya beda kurang dari 0,10 harus dibuang atau ditolak.

Indeks daya beda yang negatif menunjukkan bahwa butir soal tersebut tidak ada gunanya sama sekali dan memberikan informasi yang menyesatkan (Azwar, 2016).

Selain daya beda dan tingkat kesukaran, parameter karakteristik butir soal pilihan ganda yakni keberfungsian pengecoh. Soal pilihan ganda memiliki beberapa alternatif jawaban. Kunci jawaban yang benar hanya ada satu dari beberapa alternatif jawaban yang tersedia, sehingga alternatif jawaban selain kunci jawaban berfungsi sebagai pengecoh atau distraktor. Azwar (2016) menyatakan bahwa efektivitas pengecoh diteliti untuk melihat apakah semua pengecoh berfungsi sebagaimana mestinya. Ratnaningsih & Isfarudi (2013) menyatakan bahwa soal yang baik harus memiliki pengecoh yang relatif homogen, sehingga sulit ditebak oleh peserta tes. Kisaran indeks keberfungsian distraktor adalah antara 0% sampai dengan 100%. Kartowagiran (2012) dan Mahjabeen et al., (2017) menyatakan bahwa jika distraktor dipilih oleh lebih dari 5% peserta tes, maka distraktor tersebut dapat dikatakan berfungsi dengan baik.

Universitas Terbuka (UT), dalam proses pembelajarannya mengukur hasil belajar mahasiswa melalui beberapa aspek, salah satunya adalah melalui Ujian Akhir Semester (UAS) dalam bentuk soal pilihan berganda yang meliputi item pernyataan atau stem, dan alternatif jawaban atau option yang meliputi kunci jawaban dan pengecoh jawaban. Untuk mengetahui apakah kriteria soal yang diujikan sudah sesuai dengan yang diharapkan, atau dengan kata lain sudah baik dalam mengukur apa yang seharusnya diukur, dapat dilakukan analisis butir soal. UT melakukan analisis butir soal UAS berdasarkan teori tes klasik yang salah satu limitasinya adalah indeks karakteristik butir soal yang sangat bergantung pada kelompok peserta tes.

Dalam menyusun soal tes untuk mengukur capaian hasil belajar mahasiswa, perlu diketahui karakteristik butir soal tersebut yang terdiri atas validitas, reliabilitas, tingkat kesukaran, daya beda, dan keberfungsian distraktor. Kualitas butir soal akan tercermin dari karakteristik butir soal tersebut. Dengan adanya keterbatasan data yang dimiliki, tujuan dari tulisan ini adalah untuk menentukan: (1) sejauhmana karakteristik butir soal UAS pada mata kuliah Dasar-dasar penyuluhan pertanian ditinjau dari segi validitas, reliabilitas, dan tingkat kesukaran; dan (2) sejauhmana tingkat kualitas soal UAS pada mata kuliah Dasar-dasar Penyuluhan.

METODE

Yang menjadi subjek dalam tulisan ini adalah mata kuliah Dasar-dasar Penyuluhan Pertanian mahasiswa program studi agribisnis UT yang menempuh UAS melalui tes pilihan berganda dengan empat (4) alternatif pilihan jawaban. Data dikumpulkan dari hasil respon mahasiswa pada ujian akhir semester (UAS) masa ujian 2023.2. Kode naskah soal 87 digunakan dengan masing-masing respon jawaban tes sebanyak 45 respon jawaban. Data dikumpulkan dari sumber aplikasi pusat pengujian.

Analisis data dilakukan dengan analisis butir soal tes yang meliputi validitas, reliabilitas, tingkat kesukaran, dan daya beda dengan pendekatan kuantitatif yang mengacu pada teori tes klasik yaitu indeks karakteristik butir soal sangat bergantung pada kelompok peserta tes. Setiap butir soal dihitung indeksnya dengan bantuan program ITEMAN kemudian hasilnya dianalisis untuk menentukan kualitas butir tes berdasarkan kriteria pada Tabel 1.

Tabel 1. Kriteria tingkat kesukaran, daya beda, dan keberfungsian pengecoh

	Indeks/Koefisien	Keputusan
Tingkat Kesukaran	kurang dari 0,10 0,10 s.d. 0,29 0,30 s.d. 0,70 0,71 s.d. 0,90 lebih dari 0,90	Ditolak Direvisi Diterima Direvisi Ditolak
Daya Beda	lebih dari 0,3 0,10 s.d 0,29 kurang dari 0,10	Diterima Direvisi Ditolak
Keberfungsian Pengecoh	lebih dari 0,05 kurang dari 0,05	Diterima Direvisi

HASIL DAN PEMBAHASAN

A. Validitas Butir Soal

Validitas item soal ujian dapat diketahui dengan melihat koefisien korelasi point biserial. Data hasil perhitungan validitas item soal ujian LUHT4211 kode naskah 87 dapat dilihat pada Tabel 2. Soal yang valid dihitung dengan cara membandingkan nilai korelasi *point biserial* yang dihasilkan dengan r tabel pada taraf signifikansi 5% serta pada derajat kebebasan (DF) sama dengan $n-2$, dimana $DF = n-2 = 398 - 2 = 396$. Dengan demikian, r tabel adalah 0,1133. Item soal dikatakan valid jika $r_{pbis} > 0,1133$.

Dari sejumlah 45 butir soal UAS mata kuliah Dasar-dasar Penyuluhan Pertanian, 4 (empat) butir soal dapat dikatakan valid karena memiliki nilai validitas diatas 0.3. Nomor butir soal tersebut adalah nomor 2, 31, 34, dan 36 seperti data pada Tabel 2.

Tabel 2. Validitas Item Soal Ujian LUHT4211 Kode Naskah 87 (diterima)

No.	Nomor Item Soal	Koefisien Korelasi <i>Point Biserial</i>	Kriteria Tingkat Validitas Soal	Validitas Soal
1	2	0.390	Tinggi	Valid
2	31	0.306	Tinggi	Valid
3	34	0.396	Tinggi	Valid
4	36	0.395	Tinggi	Valid

Hal ini menunjukkan bahwa keempat butir soal yang valid tersebut adalah butir soal yang dapat mengukur apa yang seharusnya diukur, atau bisa diartikan sebagai ketepatan pengukuran yang dimiliki oleh butir soal. Butir soal tersebut dianggap memenuhi kriteria berikut, yaitu sesuai dengan kisi-kisi soal, mengacu pada indikator pembelajaran, dan hasil yang diperoleh sesuai dengan teori.

Sedangkan butir soal lainnya memiliki nilai validitas kurang dari 0.3 yang artinya memerlukan revisi soal, hasil analisisnya adalah sebagai berikut sebagai berikut seperti data yang ditampilkan pada Tabel 3.

Tabel 3. Validitas Item Soal Ujian LUHT4211 Kode Naskah 87 (revisi)

No.	Nomor Item Soal	Koefisien Korelasi <i>Point Biserial</i>	Kriteria Tingkat Validitas Soal	Validitas Soal
1	3	0.257	Cukup Baik	Valid
2	4	0.255	Cukup Baik	Valid
3	7	0.296	Cukup Baik	Valid
4	20	0.232	Cukup Baik	Valid
5	21	0.285	Cukup Baik	Valid
6	23	0.277	Cukup Baik	Valid
7	25	0.233	Cukup Baik	Valid
8	27	0.232	Cukup Baik	Valid
9	28	0.254	Cukup Baik	Valid
10	33	0.254	Cukup Baik	Valid
11	35	0.266	Cukup Baik	Valid
12	40	0.216	Cukup Baik	Valid
13	42	0.218	Cukup Baik	Valid
14	44	0.283	Cukup Baik	Valid

Dari Tabel 3 dapat diketahui bahwa item soal yang perlu direvisi adalah sebanyak 31 persen. Sebanyak 14 item soal tersebut masih bisa dikatakan valid walaupun nilainya kurang dari 0.3, namun jika butir soal tersebut masih ingin digunakan maka perlu dilakukan revisi supaya kualitasnya lebih baik.

Untuk butir soal lainnya yang memiliki status nilai validitas kurang dari 0.11 yang artinya merupakan soal yang bermasalah, dianalisis sebagai berikut seperti data yang disajikan pada Tabel 4.

Tabel 4. Validitas Item Soal Ujian LUHT4211 Kode Naskah 87 (bermasalah)

No.	Nomor Item Soal Bermasalah	Koefisien Korelasi <i>Point Biserial</i>	Kriteria Tingkat Validitas Soal	Validitas Soal
1	1	0.052	Sangat rendah	tidak valid
2	5	0.151	rendah	valid
3	6	0.141	rendah	valid
4	8	0.100	sangat rendah	tidak valid
5	9	0.155	rendah	valid
6	10	0.138	rendah	valid
7	11	0.181	rendah	valid
8	12	0.181	rendah	valid
9	13	0.080	Sangat rendah	tidak valid
10	14	0.007	Sangat rendah	tidak valid
11	15	0.132	rendah	valid
12	16	0.178	rendah	valid
13	17	0.112	rendah	valid
14	18	0.000	sangat rendah	tidak valid
15	19	0.110	rendah	valid
16	22	0.081	sangat rendah	tidak valid
17	24	0.173	rendah	valid
18	26	0.163	rendah	valid
19	29	0.165	rendah	valid
20	30	0.174	rendah	valid
21	32	0.126	rendah	valid
23	37	0.107	sangat rendah	tidak valid
24	38	0.184	rendah	valid
25	39	0.133	rendah	valid
26	41	0.017	sangat rendah	tidak valid
27	43	0.124	rendah	valid
28	45	0.132	rendah	valid

Berdasarkan data pada Tabel 4, dari 28 item soal yang bermasalah, hanya terdapat 20 item soal yang masih bisa dinyatakan valid karena memiliki nilai diatas 0.11 yaitu item soal nomor 5, 6, 9, 10, 11,12, 15,16,17,19,24,26,29,30,32,38,39,43, dan 45. Sedangkan Item soal nomor 1,8, 13,14,22,37, dan 41 dinyatakan tidak valid karena memiliki tingkat validitas yang sangat rendah. Hal ini berarti item soal tersebut kualitas kurang baik dalam mengukur hasil belajar mahasiswa. Oleh karena itu, item soal tersebut perlu direvisi, atau bahkan pada item soal yang nilai validasinya sangat rendah sebaiknya diganti dengan soal baru.

B. Reliabilitas Soal Ujian

Uji reliabilitas dilakukan dengan menggunakan K-R. 20, dimana tingkat reliabilitas soal diukur antara nilai 0 dan 1. Semakin mendekati 1 maka tingkat reliabilitas semakin tinggi. Selain itu, tingkat reliabilitas dinyatakan tinggi jika koefisien reliabilitas $> 0,7$. Hasil perhitungan menunjukkan bahwa koefisien reliabilitas soal ujian pada mata kuliah Dasar-dasar Penyuluhan adalah 0,39. Karena nilai reliabilitas tersebut masih dibawah 0.7, maka soal ujian Mata Kuliah Dasar-dasar Penyuluhan Pertanian dengan Kode Naskah 87 masih dapat dinyatakan reliabel walaupun dengan tingkat reliabilitas yang rendah.

Dengan tingkat reliabilitas yang rendah, kualitas soal ujian dapat dikatakan kurang baik. Oleh karena itu perlu ada upaya untuk perbaikan melalui proses revisi soal. Jika perlu, mengganti soal dengan yang baru agar dapat mencapai tingkat reliabilitas yang lebih baik agar mendapatkan nilai reliabilitas diatas 0.7.

C. Tingkat Kesukaran Soal Ujian

Tingkat kesukaran soal ujian adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Umumnya proporsi indeks tingkat kesukaran soal ujian berada pada rentang 0,00 – 1,00, dengan rincian sebagai berikut: 0,00-0,30 (Sukar), 0.31-0.70 (Sedang), dan 0.71-1.00 (Mudah). Sedangkan data hasil perhitungan tingkat kesukaran item soal ujian LUHT4211 kode naskah 87 dapat dilihat pada Tabel 5.

Tabel 5. Tingkat kesukaran item soal ujian LUHT4211 kode naskah 87

No.	Nomor Item Soal Bermasalah	Indeks Kesukaran (p)	Kriteria Tingkat Kesukaran Soal	Keputusan
1	1	0.281	Sukar	Revisi
2	5	0.422	Sedang	Terima
3	6	0.314	Sedang	Terima
4	8	0.822	Mudah	Revisi
5	9	0.673	Sedang	Terima
6	10	0.535	Sedang	Terima
7	11	0.618	Sedang	Terima
8	12	0.224	Sukar	Revisi
9	13	0.239	Sukar	Revisi
10	14	0.595	Sedang	Terima
11	15	0.226	Sukar	Revisi
12	16	0.136	Sukar	Revisi
13	17	0.402	Sedang	Terima
14	18	0.116	Sukar	Revisi
15	19	0	Sukar	Tolak

16	22	0.304	Sukar	Terima
17	24	0.226	Sukar	Revisi
18	26	0.244	Sukar	Revisi
19	29	0.236	Sukar	Revisi
20	30	0.314	Sedang	Terima
21	32	0.173	Sukar	Revisi
22	37	0.224	Sukar	Revisi
23	38	0.302	Sukar	Terima
24	39	0.467	Sedang	Terima
25	41	0.171	Sukar	Revisi
26	43	0.211	Sukar	Revisi
27	45	0.628	Mudah	Terima

Berdasarkan Tabel 5, dari 27 item soal yang bermasalah, terdapat 2 butir soal yang tingkat kesukarannya tergolong mudah yaitu butir nomor 8 dan butir nomor 45. Sedangkan yang tingkat kesukarannya tergolong sedang adalah sebanyak 9 item yaitu butir soal nomor 5,6,9,10,11,14,17,30, dan 39. Selanjutnya butir soal yang masuk dalam kategori sulit adalah sebanyak 16 item yaitu nomor 1,12,13,15,16,18,19,22,24,26,29,32,37,38,41, dan 43. Fakta ini menunjukkan bahwa butir soal mata kuliah Dasar-dasar Penyuluhan dominan tergolong Sukar. Hasil perhitungan ini dapat digunakan sebagai pertimbangan bagi pembuat soal untuk dapat menyesuaikan kembali tingkat kesukaran soal, baik dalam mengganti soal secara keseluruhan maupun dalam mengganti distraktor pada pilihan jawaban.

Dari data pada Tabel 5 juga bisa dilihat keputusan yang bisa diambil terkait tingkat kesukaran butir soal (mengacu pada indeks/koefisien tingkat kesukaran butir soal). Hanya 1 butir soal yang dinyatakan ditolak yaitu butir soal nomor 19. Jumlah butir soal yang diterima adalah sebanyak 12 butir soal, yaitu butir soal nomor 5,6,9,10,11,14,17,22,30,38,39,45. Sedangkan yang masih memerlukan revisi adalah sebanyak 14 butir soal yaitu nomor 1,8,12,13,15,16,18,24,26,29,32,37,41, dan 43.

Dari 27 item soal yang bermasalah, seluruhnya termasuk ke dalam klasifikasi butir soal yang buruk atau kurang baik. Artinya, seluruh item soal tersebut tidak baik dalam membedakan kelompok mahasiswa yang memiliki kemampuan tinggi (belajar) dengan kelompok mahasiswa yang memiliki kemampuan rendah (tidak belajar). Dengan demikian, seluruh item soal dengan kriteria buruk atau kurang baik pada daya pembeda direkomendasikan untuk diganti.

Tingkat validitas butir soal dengan menggunakan korelasi *point biserial* mengukur bagaimana baiknya item soal berkorelasi dengan *criterion*. Item soal dikatakan valid apabila

mempunyai dukungan yang besar terhadap skor total (Arikunto, 2018). Oleh karena itu, item soal memiliki tingkat validitas yang tinggi jika skor pada item memiliki korelasi dengan skor total. Tingkat validitas item soal ujian Mata Kuliah Dasar-dasar Penyuluhan Pertanian dengan kode naskah 87 dipengaruhi oleh 2 faktor, diantaranya: (1) kualitas soal ujian dalam mengukur evaluasi hasil belajar mahasiswa berdasarkan capaian pembelajaran; dan (2) kemampuan/pemahaman mahasiswa dalam menjawab soal ujian.

Selain kualitas soal ujian, tingkat validitas item soal ujian juga dipengaruhi kemampuan atau pemahaman mahasiswa dalam menjawab soal ujian yang diberikan. Kemampuan atau pemahaman mahasiswa tersebut diukur berdasarkan motivasi dan kinerja belajar yang tidak dijelaskan dalam hasil analisis ini. Analisis butir soal ujian menggunakan premis bahwa “seluruh peserta ujian dianggap belajar”, sedangkan masih ada probabilitas mahasiswa menebak (*guessing*) jawaban berdasarkan pilihan yang tersedia. Analisis butir soal ujian ini tidak menyertakan parameter tebakan semu (*pseudo-guessing*). Artinya, jika mahasiswa menjawab soal ujian dengan cara menebak, diduga hal ini juga berkontribusi terhadap tingkat validitas soal ujian.

SIMPULAN DAN SARAN

Berdasarkan paparan diatas, dapat disimpulkan bahwa: (1) Tingkat validitas butir soal mata kuliah Dasar-dasar Penyuluhan Pertanian tergolong rendah, sedangkan tingkat reliabilitas butir soal tergolong sedang; serta tingkat kesukaran butir soal yang masuk ke dalam kategori sedang (88%), sukar (4%), mudah (5,3%), sangat sukar (0%), dan sangat mudah (2,7%); (2) Secara umum, kualitas butir soal mata kuliah Dasar-dasar Penyuluhan tergolong kurang baik. Dari temuan tersebut, beberapa saran yang dapat disampaikan yaitu; (1) butir soal tes yang perlu direvisi diharapkan segera direvisi untuk memperbaiki kualitas butir soal yang bersangkutan; (2) butir soal yang ditolak sebaiknya tidak diikutkan dalam soal tes selanjutnya. Namun demikian, jika memungkinkan karena secara keseluruhan hasil analisis butir soal mata kuliah ini tidak bagus maka disarankan untuk mengganti set soal dengan yang baru. Selain itu, mengingat penelitian ini menganalisis butir soal dengan pendekatan teori tes klasik sehingga disarankan untuk penelitian selanjutnya dengan topik yang sama dapat ditambahkan dengan pendekatan teori respon butir.

DAFTAR PUSTAKA

Amalia, A.N. & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas XII SMA Mata Pelajaran Ekonomi Akuntansi di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, Vol. X, No. 1, 1-26.

Azwar, S. (2016). *Tes prestasi: Fungsi pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.

Iskandar, A. & Rizal, M. (2017). Analisis Kualitas Soal di Perguruan Tinggi Berbasis Aplikasi TAP. *Jurnal Penelitian dan Evaluasi Pendidikan*, Volume 21, No 2, 12-23.

Kartowagiran, B. (2012). Penulisan Butir Soal. *Makalah*. Disampaikan pada Pelatihan Penulisan dan Analisis Butir Soal bagi Sumber Daya PNS Dik-Rekinpeg, di Hotel Kawanua Aerotel, Jakarta pada tanggal 10 Oktober 2012.

Kholis, R.A.N. (2017). Analisis Tingkat Kesulitan (*Difficulty Level*) Soal pada Buku Sejarah Kebudayaan Islam Kurikulum 2013. *Jurnal Pendidikan Agama Islam*, Vol. XIV, No. 2, 305-315.

Kustati, M., & Amelia, R. (2024). Pelajaran Pai Di Sma Negeri 10 (Telaah Terhadap Reliabilitas , Daya Kesukaran Menggunakan Software Anates). 7, 6954–6961.

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 Pakistan Institute of Medical Sciences*, 13(4), 310–315.

Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan: Edisi 2*. Yogyakarta: Parama Publising.

Nuriyah, N. (2014). Evaluasi Pembelajaran: Sebuah Kajian Teori. *Jurnal Edueksos*, Vol. III, No. 1, 73-86.

Ratnaningsih, D.J. & Isfarudi. (2013). Analisis Butir Tes Objektif Ujian Akhir Semester Mahasiswa Universitas Terbuka Berdasarkan Teori Tes Modern. *Jurnal Pendidikan Terbuka dan Jarak Jauh*, Volume 14, Nomor 2, 98-109.

Sari, P., I., & Yudha, R., I. 2020. Pemanfaatan Penerapan Media Berbasis Software Anates pada Mata Kuliah Evaluasi Pembelajaran di Universitas Batanghari Jambi. *Jurnal Ilmiah Universitas Batanghari Jambi*, 20(1), 81. <https://doi.org/10.33087/jiubj.v20i1.732>