

**ANALISIS BUTIR SOAL TUGAS FAKULTAS TARBIYAH DAN ILMU
KEGURUAN PENDIDIKAN AGAMA ISLAM ISLAM UNIVERSITAS ISLAM
NEGERI K.H. ABDURRAHMAN WAHID PEKALONGAN MENGGUNAKAN
PROGRAM ANATES 4.0 FOR WINDOWS PC**

Zahra Kameliana¹, Ilham Bintang², Laela Hanifah³, Isval Maulana⁴, Mamun Hanif⁵

¹zahra.kameliana24088mhs.uingusdur.ac.id

²ilham.bintang24096@mhs.uingusdur.ac.id

³laela.hanifah24112@mhs.uingusdur.ac.id

⁴isval.maulana24115@mhs.uingusdur.ac.id

⁵mamun.hanif63@gmail.com

Tersedia Online di

<http://journal.unram.ac.id/index.php/pendas/index>

Sejarah Artikel

Diterima :

Disetujui :

Dipublikasikan :

Kata Kunci:

*Anates Program, Realbility,
Diffentiating power, Diffuculty level*

Abstack: *The purpose of this study was to determine the quality of midterm test questions for students of the Faculty of Tarbiyah and Teacher Training (FTK), Islamic Education Study Program (Pendidikan Agama Islam/PAI), using the Anates version 4.0 program. This research employed a quantitative approach with item analysis techniques, focusing on 10 essay questions. The population consisted of 60 students, with all members of the population taken as the research sample.*

The results showed that the reliability coefficient of the test items was categorized as very high (0.88). Based on the analysis of item discrimination power, only four items were considered appropriate for use, namely questions number 6, 7, 8, and 9, while the remaining items were not suitable. In terms of difficulty level, none of the items met the ideal criteria, as all questions were classified as easy to very easy, with difficulty indices

above 0.30. Furthermore, the validity analysis indicated that one item was invalid, while the other nine items were valid. Overall, the midterm test questions did not meet the standards of a good assessment instrument and were not adequate to accurately measure students' learning outcomes in Islamic Education (PAI).

Keywords: *Anates Program, Realbility, Diffentiating power, Diffuculty level*

Abstrak: Penelitian ini bertujuan untuk mengetahui kualitas soal Ujian Tengah Semester (UTS) pada mahasiswa Fakultas Tarbiyah dan Keguruan (FTK), Program Studi Pendidikan Agama Islam (PAI) Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan dengan menggunakan program Anates versi 4. Penelitian ini menggunakan metode kuantitatif dengan teknik analisis butir soal, khususnya pada 10 soal uraian yang diujikan. Populasi dalam penelitian ini berjumlah 60 mahasiswa, dengan seluruh populasi dijadikan sebagai sampel penelitian. Hasil penelitian menunjukkan bahwa tingkat reliabilitas soal memiliki kualitas sangat tinggi, yaitu sebesar 0,88. Berdasarkan analisis daya pembeda, hanya 4 soal yang dinyatakan layak digunakan, yaitu soal nomor 6, 7, 8, dan 9, sedangkan soal nomor 1, 2, 3, 4, 5, dan 10 dinyatakan tidak layak digunakan. Pada analisis tingkat kesukaran, tidak terdapat soal yang memenuhi kriteria ideal, karena seluruh soal berada pada kategori mudah dan sangat mudah dengan indeks kesukaran di atas 0,30. Sementara itu, hasil analisis validitas menunjukkan bahwa terdapat satu soal yang tidak valid, yaitu soal nomor 3, sedangkan sembilan soal lainnya dinyatakan valid. Secara umum, soal UTS pada mahasiswa Program Studi Pendidikan Agama Islam (PAI) FTK UIN K.H. Abdurrahman Wahid Pekalongan belum memenuhi standar instrumen penilaian pembelajaran yang baik dan belum dapat digunakan secara optimal untuk mengukur hasil belajar mahasiswa.

PENDAHULUAN

Penilaian pembelajaran merupakan bagian dari sistem pendidikan, sedangkan sistem pendidikan itu sendiri merupakan pelaksanaan dari program kurikulum (Oemar Hamalik, 2008). Evaluasi adalah penaksiran atau penentuan sesuatu. Proses evaluasi digunakan dalam bidang pendidikan untuk menguji kualitas proses belajar mengajar yang bertujuan agar lembaga pendidikan mengetahui hal-hal yang dapat dilakukan untuk meningkatkan kualitas pendidikan (Tomas de Aquino, 2020). Evaluasi digunakan untuk mengukur kemajuan mahasiswa, mereformasi sistem pendidikan, dan meningkatkan akuntabilitas hasil (Love, 2010). Dosen memiliki peran sebagai evaluator yang melakukan evaluasi dengan tujuan mengetahui sejauh mana keberhasilan mahasiswa dalam pembelajaran (Lufri, 2007).

Dosen memiliki peran yang tidak terlepas dari merencanakan, melaksanakan, dan mengevaluasi pembelajaran secara maksimal. Dosen melakukan evaluasi dengan tujuan mengetahui seberapa baik mahasiswa belajar serta melihat sejauh mana keberhasilan pembelajaran (Putri Maiyusri Eka, Syamsurizal Syamsurizal, 2022). Evaluasi merupakan komponen penting dalam proses belajar mengajar yang melibatkan mahasiswa dan dosen. Evaluasi memiliki peran yang sangat besar karena membantu dosen dan mahasiswa dalam mengembangkan proses pembelajaran. Evaluasi merupakan proses berkesinambungan yang membantu dalam aspek penilaian, status pendidikan, maupun pencapaian mahasiswa (Shinde, 2022; Ifeoma, 2022). Tingkat keberhasilan pendidikan sangat ditentukan oleh proses pembelajaran. Dosen berperan penting dalam pelaksanaan evaluasi karena analisis hasil evaluasi dapat menunjukkan kekuatan dan kelemahan proses pembelajaran sebagai bahan perbaikan di masa depan (Klodiana Kolomitro, 2017).

Evaluasi adalah kegiatan untuk menyimpulkan sejauh mana tujuan pembelajaran telah tercapai. Dengan demikian, kemajuan proses dan hasil belajar mahasiswa dapat ditentukan melalui penilaian. Secara umum terdapat tiga ranah penilaian, yaitu psikomotorik, afektif, dan kognitif (Arif, 2015). Evaluasi pada hakikatnya menjadi dasar dalam pengambilan keputusan, perumusan kebijakan, serta penentuan tindak lanjut pembelajaran. Kegiatan asesmen merupakan bagian integral dari setiap program pembelajaran (Suardipa & Primayana, 2020).

Evaluasi pembelajaran merupakan kegiatan yang harus dikelola secara sistematis. Evaluasi juga merupakan penyelidikan sistematis terhadap nilai suatu objek (Nurzannah et al., 2020). Kualitas pendidikan sangat dipengaruhi oleh hasil dan keakuratan penilaian (Syafri et al., 2019), sehingga hasil penilaian berpengaruh terhadap mutu pendidikan (Fariatma et al., 2017).

Tes, pengukuran, dan evaluasi merupakan tiga komponen penting dalam perencanaan pembelajaran. Perencanaan sistem pembelajaran harus disertai dengan perencanaan penilaian yang baik agar dapat mengetahui efektivitas program pembelajaran serta tingkat keberhasilan mahasiswa (Sanjaya, 2008). Evaluasi pembelajaran dilakukan melalui penyusunan instrumen,

pengumpulan data, pengolahan, hingga pemanfaatan hasil evaluasi (Arikunto, 2013). Pada umumnya, dosen beranggapan bahwa soal ujian yang disusun sudah baik, sehingga diharapkan hasil mahasiswa juga baik. Namun, jika hasil tidak sesuai harapan, maka perlu dilakukan analisis butir soal yang meliputi tingkat kesukaran, daya pembeda, distraktor, reliabilitas, dan validitas (Arif, 2015). Selain itu, penyusunan instrumen yang baik perlu memperhatikan tujuan

tes, penyusunan soal, validasi oleh ahli, serta uji coba dan analisis instrumen (Aiken, 1994 dalam Ariany & Al-Ghifari, 2018). Analisis ini penting untuk memastikan kesesuaian instrumen dengan indikator pembelajaran, metode, dan strategi pembelajaran yang digunakan (Subari et al., 2021).

Menurut Purwanto (2011), analisis soal setidaknya mencakup tingkat kesukaran, daya pembeda, efektivitas distraktor, serta validitas soal. Salah satu teknologi yang dapat digunakan dalam analisis butir soal adalah program Anates. Program ini dikembangkan oleh Drs. Karno, M.Pd., dan Yusuf Wibisono, S.T., yang mampu menghitung analisis butir soal secara cepat, tepat, dan mudah (Subari et al., 2021; Wiguna, 2021). Program Anates memiliki fitur untuk menghitung reliabilitas, daya pembeda, tingkat kesukaran, validitas, serta kualitas distraktor.

Analisis butir soal bertujuan untuk mengetahui kualitas tes secara keseluruhan maupun tiap butir soal. Hasil analisis digunakan sebagai dasar perbaikan instrumen evaluasi agar menjadi lebih baik di masa mendatang (Alpusari, 2015). Program Anates versi Windows dapat dimanfaatkan untuk menghitung berbagai indikator kualitas soal seperti reliabilitas, daya pembeda, tingkat kesukaran, dan validitas (Elviana, 2020). Instrumen tes yang baik harus memiliki validitas tinggi, reliabilitas tinggi, daya pembeda baik, dan tingkat kesukaran yang proporsional (Alpusari, 2015).

Di Fakultas Tarbiyah dan Keguruan (FTK) Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan, evaluasi pembelajaran berupa Ujian Tengah Semester (UTS) umumnya menggunakan bentuk soal uraian. Namun, analisis butir soal masih jarang dilakukan sehingga kualitas soal yang digunakan belum tentu memenuhi kriteria validitas dan reliabilitas. Oleh karena itu, penelitian ini dilakukan dengan memanfaatkan program Anates 4.0 for Windows untuk menganalisis butir soal UTS pada mahasiswa Program Studi Pendidikan Agama Islam (PAI) FTK UIN K.H. Abdurrahman Wahid Pekalongan. Penelitian ini bertujuan untuk mengetahui kualitas soal yang digunakan sehingga dapat menjadi dasar dalam penyusunan instrumen evaluasi yang lebih baik serta sebagai bank soal di masa mendatang.

METODE

Penelitian dilaksanakan di Fakultas Tarbiyah dan Keguruan (FTK) Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan pada semester genap tahun akademik 2025/2026. Subjek penelitian adalah mahasiswa pada lingkungan FTK Program Studi Pendidikan Agama Islam (PAI) dengan jumlah sampel sebanyak 60 mahasiswa. Metode yang digunakan dalam penelitian ini adalah metode deskriptif dengan pendekatan kuantitatif untuk menilai kelayakan soal yang diujikan. Soal yang dianalisis merupakan soal Ujian Tengah Semester (UTS) yang disusun oleh dosen pada salah satu mata kuliah di Program Studi Pendidikan Agama Islam (PAI). Penelitian ini menggunakan program Anates 4.0 For Windows PC untuk menganalisis reliabilitas, daya pembeda, tingkat kesukaran, dan validitas butir soal. Teknik pengumpulan data dilakukan melalui instrumen tes berupa soal uraian, lembar jawaban mahasiswa, dan kunci jawaban. Tahap persiapan dilakukan dengan menyusun kisi-kisi soal, membuat butir soal yang akan dianalisis, serta menyiapkan perangkat lunak Anates 4.0 For Windows PC. Selanjutnya, soal diujikan kepada mahasiswa dan dilakukan pemeriksaan terhadap hasil jawaban mahasiswa. Hasil jawaban tersebut kemudian dianalisis menggunakan program Anates 4.0 For Windows PC. Tahap akhir penelitian adalah melakukan analisis untuk mengetahui tingkat reliabilitas, daya pembeda, tingkat kesukaran, dan validitas butir soal. Adapun tahapan analisis butir soal adalah sebagai berikut:

a. Uji Reliabilitas

Reliabilitas merupakan ukuran yang menyatakan tingkat kekonsistenan sebuah soal. Reliabilitas berhubungan dengan apakah suatu tes dapat dipercaya sesuai dengan kriteria yang telah ditetapkan. Berikut adalah tabel kriteria reliabilitas untuk menafsirkan reliabilitas suatu soal.

Tabel 1. Kriteria Reliabilitas

Indeks Reliabilitas (r)	Tingkat Reliabilitas
$0,80 < r \leq 1,00$	Sangat tinggi
$0,60 < r \leq 0,80$	Tinggi
$0,40 < r \leq 0,60$	Sedang
$0,20 < r \leq 0,40$	Rendah
$0,00 < r \leq 0,20$	Sangat rendah

(Sumber : Arikunto, 2013)

Ketentuan :

- a. Item tersebut reliabel jika r lebih besar dari 0,60
- b. Item tersebut tidak reliabel jika r kurang dari 0,60.

b. Daya Pembeda

Daya pembeda adalah pengukuran sebuah soal untuk membedakan kemampuan mahasiswa antara yang memiliki kemampuan tinggi dengan yang memiliki kemampuan rendah. Kriteria untuk menampilkan daya pembeda terlihat pada tabel berikut.

Tabel 2. Kriteria Penafsiran

Daya Pembeda	Klasifikasi
0,00 - 0,20	Jelek (<i>poor</i>)
0,21 - 0,40	Cukup (<i>satisfactory</i>)
0,41 - 0,70	Baik (<i>good</i>)
0,71 - 1,00	Baik sekali (<i>excellent</i>)

(Sumber : Sudijono, 2018)

c. Tingkat Kesukaran

Tingkat kesukaran adalah pengukuran seberapa besar kesukaran sebuah soal. Syarat instrumen tes yang baik adalah memiliki tingkat kesukaran yang tidak terlalu sulit maupun terlalu mudah. Adapun kriteria penafsiran tingkat kesukaran berdasarkan tabel di bawah.

Tabel 3. Kriteria Penafsiran

Indeks Kesukaran	Klasifikasi
0,00 - 0,30	Sukar
0,31 - 0,70	Sedang
0,71 - 1,00	Mudah

(Sumber : Sudijono, 2018)

d. Validitas Soal

Validitas soal merupakan suatu keselarasan suatu soal dengan perangkat soal-soal lain (Syarif & Syamsurizal, 2019).

Populasi pada penelitian ini adalah seluruh mahasiswa Fakultas Tarbiyah dan Keguruan (FTK) Universitas Islam Negeri K.H. Abdurrahman Wahid Pekalongan sebanyak 60 mahasiswa dengan jumlah sampel 60 mahasiswa. Tabel berikut menampilkan jumlah mahasiswa yang termasuk dalam populasi ini.

Tabel 4. Jumlah Populasi

No	Kelas	Jumlah Siswa
1	A	32
2	B	28

Instrumen yang diteliti berupa tes uraian atau esai yang terdiri dari 10 soal. Tes uraian ini digunakan pada Ujian Tengah Semester (UTS) salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan pada semester genap tahun akademik 2025/2026. Dalam penelitian ini, digunakan pertanyaan-pertanyaan berikut.

Tabel 5. Soal dan Tingkat Kognitif Soal

No	Soal	Kognitif
1	Interaksi antar ruang berdampak pada bidang pendidikan. Apa saja dampak positif yang ditimbulkan pada bidang pendidikan tersebut.	C2
2	Deskripsikan dampak apa saja yang terjadi akibat gempa bumi dan tuliskan jenis-jenis gempa bumi yang kamu ketahui	C2
3	Salah satu bentuk interaksi antar ruang adalah perpindahan penduduk. Faktor-faktor apa saja yang menyebabkan perpindahan penduduk tersebut.	C4
4	Uraikan 5 bentuk interaksi keruangan yang ada dilingkungan kita	C2

5	Identifikasikanlah manfaat kerja sama yang terdapat pada bidang politik antar negara	C4
6	Deskripsikan pengertian konsep interaksi antar ruang	C2
7	Berdasarkan interaksi antar ruang, tuliskan bentuk interaksi dibidang sosial	C4
8	Uraikan latar belakang terjadinya kerja sama antar negara	C4
9	Bencana alam yang terjadi di negara anggota ASEAN menimbulkan dampak tidak hanya dalam satu negara. Jelaskan bagaimana interaksi negara-negara anggota ASEAN dalam menangani hal tersebut	C2
10	Singapura adalah negara kecil yang tidak memiliki kekayaan alam tetapi mampu menjadi negara maju. Faktor apa saja yang menyebabkan singapura menjadi negara maju	C4

HASIL DAN PEMBAHASAN

1. Hasil

a. Uji Reliabilitas

Hasil uji reliabilitas soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan menunjukkan nilai sebesar 0,88 berdasarkan perhitungan menggunakan program Anates 4.0 For Windows PC.

b. Daya Pembeda

Daya pembeda didapatkan dari pada pengujian soal UTS dengan Anates 4.0 For Windows PC Program dipaparkan dengan tabel di bawah.

Tabel 6. Persentase daya pembeda pada kelompok atas dan pada kelompok bawah

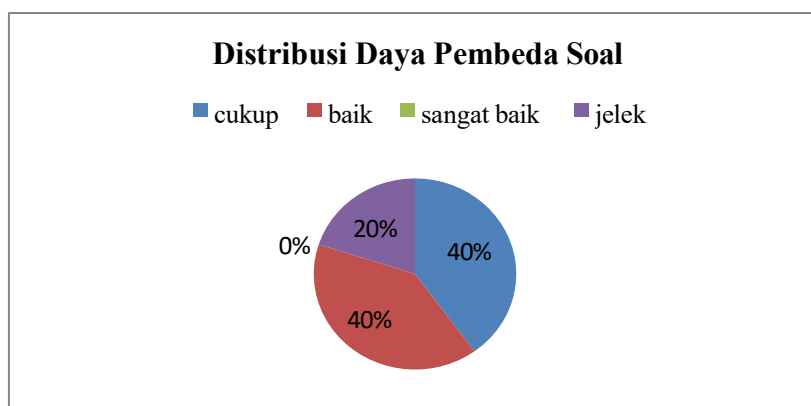
No soal	Kelompok atas	Kelompok bawah	Persentasi daya beda
1	100,00	63,44	36,56
2	96,56	77,19	19,38
3	99,06	77,50	21,56
4	97,81	74,38	23,44
5	98,13	57,50	40,63
6	96,25	51,88	44,38
7	98,75	50,00	48,75
8	97,50	52,81	44,69
9	93,44	52,19	41,25
10	98,75	80,63	18,13

(Sumber : Analisis dari Anates 4.0 For Windows PC Program)

Tabel 7. Sebaran Daya Beda

Daya beda	Nomor soal	Jumlah	Persentase
0,00 - 0,20 Jelek (poor)	2, 10	2	20 %
0,21 - 0,40 Cukup (satisfactory)	1, 3, 4, 5	4	40%
0,41 - 0,70 Baik (good)	6,7,8,9	4	40%
0,71 - 1,00 Baik sekali (excellent)	-	-	-

(Sumber : Analisis dari Anates 4.0 For Windows PC Program)



Gambar 1. Sebaran Daya Beda Soal Ujian Tengah Semester (UTS) pada Mahasiswa FTK UIN K.H. Abdurrahman Wahid Pekalongan

Berdasarkan Tabel 6, daya pembeda soal yang dihitung menggunakan program Anates 4.0 For Windows PC diperoleh melalui perbandingan antara kelompok atas dan kelompok bawah. Pembagian kelompok ini bertujuan untuk melihat kemampuan soal dalam membedakan mahasiswa yang memiliki tingkat pemahaman tinggi dengan mahasiswa yang memiliki tingkat pemahaman rendah. Hasil analisis menunjukkan bahwa daya pembeda soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan diperoleh dari 10 butir soal uraian yang telah diujikan kepada mahasiswa. Setiap butir soal memiliki kemampuan yang berbeda dalam membedakan tingkat penguasaan materi mahasiswa..

c. Tingkat Kesukaran

Tabel 8. Hasil Tingkat Kesukaran pada Mahasiswa FTK UIN K.H. Abdurrahman Wahid Pekalongan

No soal	Kesukaran %	Tafsiran
1	81,72	Mudah
2	86,88	Sangat Mudah
3	88,28	Sangat Mudah
4	86,09	Sangat Mudah
5	77,81	Mudah
6	74,06	Mudah
7	74,38	Mudah
8	75,16	Mudah
9	72,81	Mudah
10	89,69	Sangat Mudah

(Sumber : Analisis dari Anates 4.0 For Windows PC Program)

Dari hasil pengolahan butir soal dengan memakai *Anates* bisa diketahui bahwa tingkatan kesulitan pada soal dapat menafsirkan tingkatan kesukaran soal tersebut. Hasil pada tingkat kesukaran masing-masing nomor soal pada indek diatas 0.00% hingga 100%.

d. Validitas Soal

Pada soal UTS yang telah dianalisis dengan *Anates 4.0 For Windows PC Program* dapat dilihat dengan tabel di bawah ini:

Tabel 9. Hasil Validitas soal UTS Mahasiswa FTK UIN K.H.

Abdurrahman Wahid Pekalongan			
No Soal	Korelasi	Signifikansi	Validitas
1	0,618	Signifikan	Valid
2	0,584	Signifikan	Valid
3	0,575	-	Tidak Valid
4	0,593	Signifikan	Valid
5	0,790	Sangat Signifikan	Valid
6	0,741	Sangat Signifikan	Valid
7	0,732	Sangat Signifikan	Valid
8	0,787	Sangat Signifikan	Valid
9	0,614	Signifikan	Valid
10	0,635	Signifikan	Valid

(Sumber : Analisis dari Anates 4.0 For Windows PC Program)

Tabel 9 memaparkan hasil uji validitas soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan. Hasil analisis menunjukkan bahwa terdapat 9 butir soal yang valid dan 1 butir soal yang tidak valid karena nilai korelasinya berada di bawah 0,580. Analisis tersebut dilakukan menggunakan program Anates 4.0 For Windows PC.

2. Pembahasan

Tindakan menganalisis butir soal merupakan langkah penting yang harus dilakukan oleh seorang dosen untuk meningkatkan kualitas instrumen evaluasi yang akan dikerjakan oleh mahasiswa. Proses ini meliputi penyusunan soal, pelaksanaan tes, perbaikan, serta evaluasi terhadap soal yang telah digunakan. Analisis ini bertujuan untuk menelaah setiap butir soal agar memiliki kualitas yang baik sebelum digunakan dalam proses penilaian. Dosen dapat melakukan analisis butir soal secara kuantitatif maupun kualitatif. Analisis secara kuantitatif dapat dilakukan dengan menguji reliabilitas, validitas, tingkat kesukaran, dan daya pembeda soal. Melalui analisis tersebut, dapat diketahui kualitas setiap butir soal secara lebih objektif. Penelitian ini bertujuan untuk mengungkap kualitas instrumen tes yang digunakan dengan melihat aspek reliabilitas, daya pembeda, tingkat kesukaran, dan validitas soal, sehingga dapat menghasilkan instrumen evaluasi yang lebih baik dan sesuai standar penilaian pembelajaran di perguruan tinggi.

a. Uji Reliabilitas

Uji Reliabilitas menurut Hasan (2014) berfungsi untuk memastikan apakah alat ukur dapat diandalkan dan mempertahankan konsistensinya bahkan setelah dilakukan pengukuran berulang kali. Instrumen yang andal adalah instrumen yang mampu mengungkapkan data yang dapat dipercaya (Syamsurizal, 2020). Hasil uji reliabilitas menunjukkan nilai sebesar 0,88 berdasarkan perhitungan menggunakan program Anates 4.0 For Windows PC. Hasil tersebut menunjukkan bahwa instrumen tes Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan memiliki tingkat keandalan yang sangat tinggi, yaitu berada pada rentang 0,80 sampai dengan 1,00.

Reliabilitas yang tinggi menunjukkan bahwa instrumen tes memiliki konsistensi yang baik dalam mengukur kemampuan mahasiswa secara stabil. Dengan demikian, tes tersebut dapat dikatakan dapat dipercaya sebagai alat ukur hasil belajar (Loka Son, 2019). Reliabilitas menunjukkan bahwa suatu instrumen cukup dapat dipercaya untuk digunakan sebagai alat pengumpul data (Syahril, Hadiyanto, Nelfia Adi, 2020). Instrumen yang reliabel akan memberikan hasil yang relatif konsisten apabila diuji pada kelompok yang sama dalam waktu yang berbeda. Berdasarkan hasil analisis, diperoleh nilai reliabilitas sebesar 0,88 pada

instrumen tes Ujian Tengah Semester (UTS) yang digunakan pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan.

b. Daya Pembeda

Ukuran daya pembeda suatu item untuk membedakan antara mahasiswa yang telah menguasai keterampilan dari mereka yang belum menguasai dapat diketahui dengan melihat ukuran indeks diskriminasi butir soal (Rahayu, R., & Djazari, 2016).

Daya pembeda yang dihitung menggunakan program Anates 4.0 For Windows PC diperoleh dengan membandingkan kelompok atas dan kelompok bawah sebagaimana disajikan pada Tabel 6. Perbandingan ini bertujuan untuk melihat kemampuan butir soal dalam membedakan mahasiswa yang memiliki tingkat pemahaman tinggi dan rendah. Berdasarkan hasil analisis, daya pembeda soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) diperoleh dari 10 butir soal uraian yang telah diujikan kepada mahasiswa.

Berdasarkan hasil analisis perhitungan dengan menggunakan Anates 4.0 For Windows PC Program didapatkan 2 nomor pada soal (20%) memiliki daya pembeda jelek dengan persentasi daya pembeda antara 0.00-0.20 yaitu pada soal ke 2 dan ke 10, 4 buah soal (40 %) memiliki daya beda cukup dengan persentasi daya beda berkisar antara 0.21-0.40, terdapat pada soal ke 1,3,4,5 dan 4 butir soal (40 %) memiliki daya beda baik dengan persentasi daya beda antara 0.41-0.70 yang terdapat pada soal ke 6,7,8,9 dan tidak ada soal (0 %) dengan daya beda pada kategori baik sekali dengan persentasi daya beda antara 0.71-1.00.

Hasil analisis daya pembeda menggunakan program Anates 4.0 For Windows PC menunjukkan bahwa dari 10 butir soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan, hanya 4 butir soal yang layak digunakan, yaitu soal nomor 6, 7, 8, dan 9. Keempat soal tersebut memiliki daya pembeda yang baik dengan rentang indeks antara 0,41–0,70. Sementara itu, 6 butir soal lainnya, yaitu soal nomor 1, 2, 3, 4, 5, dan 10, dinyatakan tidak layak digunakan karena memiliki daya pembeda rendah hingga cukup, dengan rentang indeks antara 0,00–0,40. Soal dengan daya pembeda rendah menunjukkan bahwa butir soal tersebut belum mampu membedakan mahasiswa yang memiliki tingkat pemahaman tinggi dengan mahasiswa yang memiliki tingkat pemahaman rendah.

Oleh karena itu, butir soal yang termasuk dalam kategori daya pembeda rendah dan cukup perlu dilakukan revisi agar kualitasnya meningkat dan dapat memenuhi kriteria sebagai soal yang baik atau sangat baik. Nomor soal memiliki daya beda pada kategori baik sekali, kategori baik dan kategori cukup seharusnya digunakan dan dimasukkan kedalam kumpulan soal, dan nomor soal yang daya beda buruk seharusnya dilakukan tindakan perbaikan soal dan mencari tau apa yang menyebabkan soal mempunyai daya beda dengan kategori jelek. Tujuan memperbaiki soal yang mempunyai daya beda dengan kategori jelek untuk bisa memberikan kejelasan bagi mahasiswa yang memiliki kompetensi tinggi dan kompetensi rendah dalam menjawab soal. Nomor soal seharusnya dapat menggambarkan terdapatnya perbedaan pemahaman diantara mahasiswa yang paham pada pembahasan terhadap mahasiswa yang belum memahami penjelasan (Muluki, 2020).

c. Tingkat Kesukaran

Pada pengukuran tingkat kesukaran pada nomor soal dapat menggambarkan tingkat kesulitan pada soal tes. Nomor soal tes dikategorikan baik apabila hasil tingkatan kesulitan nomor soal pada kategori sedang atau soal dengan kategori tidak terlalu mudah dan tidak terlalu susah (Fietri et al., 2021)

Tingkat kesukaran pada soal dikatakan layak apabila soal tidak termasuk dalam kategori susah dan mudah. Pada soal mudah maka soal tersebut tidak membuat mahasiswa meningkatkan usaha menyelesaikannya, sedangkan ketika soal sangat sulit maka akan membuat mahasiswa menyerah dalam mengerjakannya dan akan menurunkan semangat

mahasiswa dalam berusaha menyelesaikannya karena soal tersebut diluar batas kemampuannya (Mujiyanto, 2017).

Hasil pengolahan butir soal dengan memakai Anates 4.0 For Windows PC Program diketahui bahwa tingkatan kesulitan pada soal dapat menafsirkan tingkatan kesukaran soal tersebut. Hasil pada tingkat kesukaran masing-masing nomor soal pada indeks diatas 0.00% hingga 100%. Pada bilangan 0.00% menunjukan bahwa nomor soal tersebut tergolong dalam klasifikasi sangat sukar, karena seluruh mahasiswa tidak dapat menjawab dengan benar soal tersebut. Dan jika angka pada indeks kesukaran menunjukkan angka sebesar 100% berarti soal masuk katagori mudah, disebabkan semua mahasiswa bisa dengan benar menjawab soal tersebut.

Pentingnya melihat tingkat kesukaran pada soal untuk memberikan dan memfasilitasi alat penilaian kesulitan belajar mahasiswa maupun dalam rangka agar dapat memberikan peningkatan pada penilaian berbasis kelas. Dari analisis soal pada tingkat kesukaran akan dapat menentukan baik tidaknya nomor soal tersebut (Fietri et al., 2021)

Butir soal bisa dikategorikan baik bilamana soal itu tidak tergolong sangat mudah atau tidak tergolong sangat sulit. Soal dengan kategori terlalu mudah tidak akan memotivasi mahasiswa untuk meningkatkan upaya penyelesaiannya dan soal yang tergolong sangat sulit dapat membuat mahasiswa menyerah dan tidak mau untuk mengulangi lagi karena berada di atas kemampuannya (Iskandar, A., & Rizal, 2018).

Hasil analisis tingkat kesukaran menunjukkan bahwa dari 10 butir soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan, seluruh soal berada pada kategori mudah dan sangat mudah. Indeks kesukaran soal berada pada rentang 72,81 sampai dengan 89,69. Terdapat 4 butir soal yang termasuk dalam kategori sangat mudah, yaitu soal nomor 2, 3, 4, dan 10 dengan indeks kesukaran antara 86,88 sampai dengan 89,69. Sementara itu, 6 butir soal lainnya termasuk dalam kategori mudah, yaitu soal nomor 1, 5, 6, 7, 8, dan 9 dengan indeks kesukaran antara 72,81 sampai dengan 81,72.

Berdasarkan hasil perhitungan analisis tingkat kesukaran menggunakan program Anates versi 4 for Windows, dari 10 butir soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK), tidak terdapat soal yang memenuhi kriteria tingkat kesukaran yang baik. Seluruh butir soal berada pada kategori mudah hingga sangat mudah, dengan persentase tingkat kesukaran berada di atas 0,31. Hal ini menunjukkan bahwa soal yang diujikan belum mampu mengukur kemampuan mahasiswa secara optimal. Oleh karena itu, sangat disarankan agar butir soal direvisi sehingga memiliki variasi tingkat kesukaran sesuai dengan kategori yang ditetapkan..

Suatu soal dikatakan baik apabila memiliki variasi tingkat kesukaran. Berdasarkan hasil analisis tingkat kesukaran menggunakan program Anates 4.0 For Windows PC, dari 10 butir soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK), seluruh soal berada pada kategori mudah. Hal ini menunjukkan bahwa soal belum mampu mendorong mahasiswa untuk berpikir lebih kritis dan meningkatkan usaha dalam menyelesaikan permasalahan. Oleh karena itu, butir soal perlu dievaluasi dan direvisi agar memiliki tingkat kesukaran minimal sedang hingga sulit, sehingga dapat mengasah kemampuan berpikir kritis dan keterampilan pemecahan masalah mahasiswa.

d. Validitas Soal

Validitas soal merupakan suatu keselarasan suatu soal dengan perangkat soal-soal lain (Syarif & Syamsurizal, 2019). Dalam mengukur validnya soal, bisa dipakai melalui teknik analisis yaitu teknik korelasi. Soal yang bisa digunakan sebagai tes untuk menentukan hasil belajar ialah soal dengan kategori valid dan soal dengan kategori tidak valid maka tidak digunakan pada tes hasil belajar (Alpusari, 2015).

Tabel 8 menunjukkan hasil analisis validitas butir soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK). Berdasarkan hasil analisis menggunakan program Anates 4.0 For Windows PC, terdapat 9 butir soal yang dinyatakan valid dan 1 butir soal yang tidak valid. Soal yang termasuk kategori valid

merupakan butir soal yang mampu mengukur apa yang seharusnya diukur secara tepat dan efektif. Dari 9 soal yang valid, sebanyak 4 butir soal memiliki tingkat validitas tinggi dan 5 butir soal memiliki tingkat validitas sedang.

Berdasarkan hasil analisis, butir soal nomor 3 dinyatakan tidak valid sehingga tidak dapat digunakan sebagai instrumen tes hasil belajar mahasiswa karena memiliki nilai korelasi di bawah 0,580. Sementara itu, butir soal yang dinyatakan valid adalah soal yang memiliki nilai korelasi di atas 0,580. Adapun jumlah soal yang dapat digunakan sebanyak 9 butir, yaitu soal nomor 1, 2, 4, 5, 6, 7, 8, 9, dan 10.

Pada dasarnya pendidikan memerlukan sebuah evaluasi. Tak hanya hasil belajar mahasiswa saja yang dievaluasi, melainkan juga alat pengukur yang digunakan untuk mengukur keberhasilan mahasiswa. Dalam suatu evaluasi pembelajaran dibutuhkan juga evaluasi proses penilaian dan pengukuran mahasiswa. Salah satunya dengan menganalisis reabilitas, daya pembeda, kesukaran soal, dan validitas soal. Untuk mengetahui apakah butir-butir tersebut sudah menjalankan fungsinya sebagaimana mestinya, maka perlu identifikasi. Dengan harapan akan menghasilkan berbagai informasi berharga, yang menjadi umpan balik guna melakukan perbaikan, pembenahan dan penyempurnaan kembali, hingga nantinya benar-benar dapat menjalankan fungsinya sebagai alat pengukur hasil belajar yang berkualitas tinggi.

KESIMPULAN

Berdasarkan hasil analisis soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) menggunakan program Anates 4.0 For Windows PC, diperoleh kesimpulan dari 10 butir soal uraian sebagai berikut:

1. Hasil analisis uji reabilitas butir soal uraian mempunyai reliabilitas soal yang sangat tinggi adalah 0.88.
2. Hasil analisis daya pembeda soal hanya 4 soal yang layak digunakan yaitu pada soal no 6,7,8,9 dengan persentasi daya pembeda antara 0.41-0.70. Sedangkan 6 butir soal 1,2,3,4, 5 dan 10 tidak layak digunakan karena memiliki daya pembeda jelek dan cukup dengan persentasi daya pembeda antara 0.00-0.40.
3. Hasil analisis tingkat kesukaran didapatkan soal tidak ada yang memenuhi kriteria tingkat kesukaran karena semua yang diujikan berada pada tingkat mudah dan sangat mudah dengan persentasi tingkat kesukaran semua berada diatas 0,31.
4. Hasil analisis validitas tes terdapat satu soal tidak valid untuk digunakan yaitu soal no 3, dan terdapat 9 butir soal yang valid untuk digunakan.

Berdasarkan paparan kesimpulan tersebut, dapat disimpulkan bahwa soal Ujian Tengah Semester (UTS) pada salah satu mata kuliah di Fakultas Tarbiyah dan Keguruan (FTK) UIN K.H. Abdurrahman Wahid Pekalongan secara umum masih belum memenuhi karakteristik sebagai instrumen evaluasi pembelajaran yang baik. Soal tersebut belum sepenuhnya layak digunakan untuk mengukur keberhasilan proses belajar mengajar, terutama ditinjau dari aspek daya pembeda dan tingkat kesukaran soal.

DAFTAR PUSTAKA

- Alpusari, M. (2015). Analisis Butir Soal Konsep Dasar Ipa 1 Melalui Penggunaan Program Komputer Anates Versi 4.0 for Windows. *Primary: Jurnal Pendidikan Guru Sekolah Dasar*, 3(2), 106. <https://doi.org/10.33578/jpfkip.v3i2.2501>
- Ariany, R. L., & Al-Ghifari, A. (2018). Penggunaan Software Anates Untuk Validasi Instrumen Tes. *Al-Khidmat*, 1(1), 73–78. <https://doi.org/10.15575/jak.v1i1.3327>
- Arif, M. (2015). Penerapan Aplikasi Anates Bentuk Soal Pilihan Ganda. *EduTic - Scientific Journal of Informatics Education*, 1(1), 1–9. <https://doi.org/10.21107/edutic.v1i1.398>
- Arikunto, S. (2013). *Prosedur Penelitian Suatu Pendekatan Praktik*. PT. Rineka Cipta.
- Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam

- Menggunakan Program Anates. *Jurnal MUDARRISUNA*, 10(2), 58–74.
- Fietri, W. A., Zulyusri, & Violita. (2021). Analisis Butir Soal Biologi Kelas XI Madrasah Aliyah Sakinah Kerinci Menggunakan Program Komputer Anates 4.0 For Windows. *Natural Science: Jurnal Penelitian Bidang IPA Dan Pendidikan IPA*, 7(1), 2477–6181.
- Friatma, A., Syamsurizal, & Helendra. (2017). Analisis Kualitas Soal Ujian Akhir Semester Genap Mata Pelajaran Biologi Kelas XI IPA SMA Negeri Wilayah Selatan Kabupaten Solok Tahun Pelajaran 2015/2016. *Bioeducation Journal*, 1(2), 50–67.
- Hasan, M. & I. (2014). *Analisis Data Penelitian Dengan Statistik* (2nd ed.). Bumi Aksara.
- Ifeoma, E. F. (2022). The Role of Evaluation in Teaching and Learning Process in Education. *International Journal of Advanced Academic and Educational Research*, 13(5), 120–129. <https://doi.org/2726145223713511>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 12–23.
- Klodiana Kolomitro. (2017). A survey on evaluation practices in teaching and learning centres. *International Journal For Academic Development*, 3(22), 186–198.
- Loka Son, A. (2019). Instrumentasi Kemampuan Pemecahan Masalah Matematis: Analisis Reliabilitas, Validitas, Tingkat Kesukaran Dan Daya Beda Butir Soal. *Gema Wiralodra*, 1(10), 41–52.
- Love, A. J. (2010). *Understanding Approaches to Evaluation* (Third Edit). International Encyclopedia of Education.
- Lufri. (2007). *Strategi Pembelajaran Biologi. Teori, Praktek dan Penelitian* (T. E. U. Press (ed.)).
- Mujianto, S. (2017). Analisis daya beda soal. taraf kesukaran, butir tes, validitas butir tes, interpretasi hasil tes valliditas ramalan dalam evaluasi pendidikan. *Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 192–213.
- Muluki, A. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(10).
- Nurzannah, N., Ginting, N., & Setiawan, H. R. (2020). Implementation Of Integrated Quality Management In The Islamic Education System. *Proceeding International Seminar of Islamic Studies*, 1(1), 1–9. <http://jurnal.umsu.ac.id/index.php/insis/article/view/4065>
- Oemar Hamalik. (2008). *Proses Belajar Mengajar* (8th ed.). PT. Bumi Aksara.
- Purwanto. (2011). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Remaja Rosdakarya.
- Putri Maiyusri Eka, Syamsurizal Syamsurizal, L. L. (2022). Analyzed Question Quality of School Examinations Class IX Science Subjects. 6(1), 40–47.
- Rahayu, R., & Djazari, M. (2016). Analisis Kualitas Soal Pra Ujian Nasional Mata Pelajaran Ekonomi Akuntansi. *Jurnal Pendidikan Akuntansi Indonesia*, 1(14).
- Sanjaya, W. (2008). *Perencanaan dan Desain Sistem Pembelajaran*. Kencana Prenada Media Group.
- Shinde, S. (2022). IMPORTANCE OF EVALUATION IN TEACHING LEARNING PROCESS. *Scholarly Research Journal for Humanity Science & English Language*,. <https://oaji.net/articles/2022/1201-1670914970.pdf>
- Suardipa, I. P., & Primayana, K. H. (2020). Peran desain evaluasi pembelajaran untuk meningkatkan kualitas pembelajaran. *Widyacarya*, 4(2), 88–100. <http://jurnal.stahnmpukuturan.ac.id/index.php/widyacarya/article/view/796>
- Subari, A., Lufri, L., & Syamsurizal, S. (2021). Analisis Butir Soal Ujian Akhir Semester Ganjil Biologi Kelas XI MAN 2 Kota Jambi. *Bioscientist : Jurnal Ilmiah Biologi*, 9(1), 45. <https://doi.org/10.33394/bjib.v9i1.3612>
- Sudijono, A. (2018). *Pengantar Evaluasi Pendidikan*. Pendidikan. Rajawali Pers.

- Syafti, N. K., Darussyamsu, R., Selaras, G. H., & Syamsurizal, S. (2019). Analyzed Quality Question Type Multiple Choices Questions (MCQ) of Final Exam Even Semester on Biology Course Class XI SMA Negeri Se-Kecamatan IV Jurai Academic Year 2017/2018. *Jurnal Atrium Pendidikan Biologi*, 4(1), 200. <https://doi.org/10.24036/apb.v4i1.5475>
- Syahril, Hadiyanto, Nelfia Adi, T. A. N. (2020). Pemberdayaan Aset Sekolah Dan Instansi Pemerintah Melalui Workshop Sistem Manajemen Aset. *Jurnal Halaqah.*, 2(2), 3.
- Syamsurizal. (2020). Validitas dan Reliabilitas Alat Ukur. *Jurnal Osf*, 1–11.
- Syarif, E. A., & Syamsurizal, S. (2019). Analyzed Quality of Senior High School Biology Olympiad Questions at West Sumatera, Riau, Jambi, and Bengkulu in 2018. *Bioeducation Journal*, 3(2), 142–150. <https://doi.org/10.24036/bioedu.v3i2.194>
- Tomas de Aquino. (2020). (Pdf) Assessment and Evaluation in Education. *Research Gate*, July. https://www.researchgate.net/publication/342918149_ASSESSMENT_AND_EVALUATION_IN_EDUCATION
- Wiguna, S. (2021). Aplikasi Anates Dalam Evaluasi Pembelajaran. *Suparyanto Dan Rosad* (2015, 5(3), 248–253.