DARI SKOR KE KUALITAS SOAL: ANALISIS BUTIR TES BERBASIS ANATES DALAM MENGUNGKAP MUTU EVALUASI PEMBELAJARAN DI PONDOK PESANTREN NURUL-HUDA PANDANSARI

Mifta Ayu Fadillah[1], Rahma Aulia Zahra[2] , Nailil Izza[3], Niswah Atqiya Maulidya[4] , M. Wildan Ahfadh[5]
Universitas Islam Negeri KH Abdurrahman Wahid Pekalongan

mifta.ayu.fadillah24124@mhs.uingusdur.ac.id[1],
rahma.aulia.zahra24132@mhs.uingusdur.ac.id[2]
nailil.izza24140@mhs.uingusdur.ac.id[3],
niswah.atqiya.maulidya24146@mhs.uingusdur.ac.id[4],
m.wildan.ahfadh24152@mhs.uingusdur.ac.id[5]

ABSTRACT

This study aims to analyze the quality of multiple-choice test items based on the Anates application in the evaluation of learning at the PDF Nurul Huda Pandansari Islamic Boarding School. This study employed a quantitative descriptive method with 20 students as subjects and an instrument consisting of 20 multiple-choice items. Data were analyzed using the Anates application to determine item validity, discriminative power, difficulty level, distractor effectiveness, and test reliability. The analysis results indicate that the test instrument has high reliability at 0.87. Regarding validity, 9 items demonstrated good validity, 8 items were not yet significant, and 3 items could not have their correlation values calculated. In terms of discriminative power, several items effectively distinguished students' abilities, particularly items 12, 15, 16, and 18, while others still require improvement. The difficulty levels of the items also varied, ranging from very difficult to very easy. Additionally, the quality of distractors in some items functioned well, but there were still some ineffective distractors. Thus, item analysis based on Anates can assist educators in assessing the quality of evaluation instruments and serve as a basis for improving test items to make them more valid, objective, and appropriate for students' abilities.

***Keywords**: item analysis, Anates, learning assessment, discriminative power, distractors, validity.*

ABSTRAK

Penelitian ini bertujuan untuk menganalisis kualitas butir soal tes pilihan ganda berbasis aplikasi Anates dalam evaluasi pembelajaran di Pondok Pesantren PDF Nurul Huda Pandansari. Penelitian ini menggunakan metode deskriptif kuantitatif

dengan subjek sebanyak 20 peserta didik dan instrumen berupa 20 butir soal pilihan ganda. Data dianalisis melalui aplikasi Anates untuk mengetahui validitas butir, daya pembeda, tingkat kesukaran, efektivitas distraktor, dan reliabilitas tes. Hasil analisis menunjukkan bahwa instrumen tes memiliki reliabilitas tinggi sebesar 0,87. Dari aspek validitas, terdapat 9 butir soal yang memiliki validitas baik, 8 butir belum signifikan, dan 3 butir tidak dapat dihitung nilai korelasinya. Dari aspek daya pembeda, beberapa butir soal mampu membedakan kemampuan peserta didik dengan baik, terutama butir nomor 12, 15, 16, dan 18, sedangkan beberapa butir lainnya masih perlu diperbaiki. Tingkat kesukaran soal juga bervariasi, mulai dari sangat sukar hingga sangat mudah. Selain itu, kualitas distraktor pada beberapa butir sudah berfungsi dengan baik, tetapi masih terdapat pengecoh yang kurang efektif. Dengan demikian, analisis butir soal berbasis Anates dapat membantu pendidik dalam menilai mutu instrumen evaluasi serta menjadi dasar perbaikan soal agar lebih valid, objektif, dan sesuai dengan kemampuan peserta didik.

Kata kunci: analisis butir soal, Anates, evaluasi pembelajaran, daya pembeda, distraktor, validitas.

A. Pendahuluan

Tes, pengukuran, dan evaluasi adalah konsep yang dig unakan dalam pendidikan untuk menjelaskan bagaimana kemajuan pembelajaran dan hasil akhir pembelajaran siswa dinilai (Adom et al., 2020). Melalui evaluasi, pendidik dapat memperoleh gambaran mengenai kemampuan peserta didik setelah mengikuti proses pembelajaran. Namun, evaluasi yang baik tidak hanya ditentukan oleh pelaksanaan tes semata, tetapi juga oleh kualitas instrumen tes yang digunakan. Instrumen tes yang kurang baik dapat menghasilkan data yang kurang tepat, sehingga penilaian terhadap kemampuan peserta didik menjadi kurang objektif (Arianda et al., 2024).

Dalam kegiatan pembelajaran, soal tes sering digunakan sebagai alat ukur untuk mengetahui hasil belajar peserta didik (Adom et al., 2020). Akan tetapi, tidak semua soal yang digunakan dalam tes memiliki kualitas yang sama. Ada soal yang mampu mengukur kemampuan peserta didik dengan baik, tetapi ada pula soal yang terlalu mudah, terlalu sulit, atau kurang mampu membedakan tingkat penguasaan peserta didik (Bulala, 2023). Oleh karena itu, analisis butir soal perlu dilakukan agar pendidik dapat mengetahui kelayakan soal yang telah digunakan dalam evaluasi pembelajaran.

Penelitian ini dilaksanakan di Pondok Pesantren PDF (Pendidikan Diniyah Formal) Nurul Huda Pandansari pada hari Kamis, 7 Mei 2026. Penelitian ini menggunakan hasil pengolahan data tes melalui aplikasi Anates sebagai dasar untuk menelaah kualitas butir soal. Data yang dianalisis terdiri atas 20 peserta didik dan 20 butir soal pilihan ganda, dengan ketentuan jawaban benar diberi skor 5 dan jawaban salah diberi skor 0. Hasil pengolahan menunjukkan bahwa skor peserta didik bervariasi, dengan skor tertinggi sebesar 80 dan skor terendah sebesar 40. Selain itu, hasil analisis menunjukkan rata-rata skor sebesar 13,45, simpang baku 2,93, serta reliabilitas tes sebesar 0,87, yang menunjukkan bahwa instrumen tes memiliki tingkat konsistensi yang tinggi.

Penggunaan aplikasi Anates dalam penelitian ini membantu proses analisis soal menjadi lebih sistematis dan terarah. Melalui hasil analisis tersebut, pendidik tidak hanya memperoleh informasi berupa skor peserta didik, tetapi juga dapat mengetahui mutu soal yang digunakan sebagai alat evaluasi (Nabilah et al., 2025). Analisis ini mencakup beberapa aspek penting, seperti ketepatan butir soal dalam mengukur kemampuan peserta didik, kemampuan soal dalam membedakan peserta didik berkemampuan tinggi dan rendah, tingkat kesukaran soal, serta keberfungsian pilihan jawaban pengecoh. Dengan demikian, hasil tes tidak hanya dipahami sebagai angka atau nilai akhir, melainkan juga sebagai bahan refleksi untuk memperbaiki kualitas penyusunan soal pada kegiatan evaluasi berikutnya.

Berdasarkan latar belakang tersebut, artikel ini disusun dengan judul “Dari Skor ke Kualitas Soal: Analisis Butir Tes Berbasis Anates dalam Mengungkap Mutu Evaluasi Pembelajaran. Artikel ini bertujuan untuk memberikan gambaran mengenai pentingnya analisis butir tes sebagai upaya meningkatkan kualitas evaluasi pembelajaran di lingkungan Pondok Pesantren PDF Nurul Huda Pandansari.

B. Metode Penelitian

Penelitian ini menggunakan metode deskriptif kuantitatif dengan pendekatan analisis butir soal berbasis aplikasi Anates. Metode deskriptif kuantitatif digunakan untuk

menggambarkan kualitas instrumen tes berdasarkan data hasil jawaban peserta didik. Melalui pendekatan ini, setiap butir soal dianalisis secara sistematis untuk mengetahui mutu soal yang digunakan dalam evaluasi pembelajaran.

Penelitian ini dilaksanakan di Pondok Pesantren PDF (Pendidikan Diniyah Formal) Nurul Huda Pandansari pada hari Kamis, 7 Mei 2026. Subjek penelitian terdiri atas 20 peserta didik yang mengikuti tes evaluasi pembelajaran. Instrumen yang dianalisis berupa 20 butir soal pilihan ganda. Dalam penskoran, jawaban benar diberi skor 5, sedangkan jawaban salah diberi skor 0. Hasil pengolahan menunjukkan bahwa skor peserta didik bervariasi, dengan skor tertinggi 80 dan skor terendah 40. Selain itu, rata-rata skor peserta didik sebesar 13,45, simpang baku 2,93, dan reliabilitas tes sebesar 0,87, sehingga tes menunjukkan tingkat konsistensi yang tinggi.

Tahapan penelitian diawali dengan persiapan instrumen tes. Pada tahap ini, peneliti menyiapkan soal pilihan ganda yang digunakan sebagai alat evaluasi pembelajaran. Selain itu, peneliti juga menyiapkan lembar jawaban serta menentukan pedoman penskoran agar proses pemeriksaan dan pengolahan data dapat dilakukan secara terarah.

Tahap berikutnya adalah pelaksanaan tes. Tes diberikan kepada peserta didik Pondok Pesantren PDF Nurul Huda Pandansari sesuai dengan waktu yang telah ditentukan. Peserta didik diminta mengerjakan soal secara mandiri. Setelah seluruh peserta didik menyelesaikan tes, lembar jawaban dikumpulkan untuk diperiksa dan diolah lebih lanjut.

Setelah data terkumpul, tahap selanjutnya adalah penginputan data ke dalam aplikasi Anates. Jawaban peserta didik dimasukkan ke dalam aplikasi tersebut untuk memperoleh hasil analisis butir soal. Melalui Anates, diperoleh informasi mengenai skor peserta didik, kelompok unggul dan kelompok asor, reliabilitas tes, validitas butir, daya pembeda, tingkat kesukaran, serta kualitas pengecoh atau distraktor.

Tahap selanjutnya adalah analisis kualitas butir soal. Analisis ini mencakup empat komponen utama, yaitu validitas, daya pembeda, tingkat kesukaran, dan distraktor. Daya pembeda digunakan untuk melihat kemampuan soal dalam membedakan

peserta didik yang memiliki kemampuan tinggi dan rendah. Tingkat kesukaran digunakan untuk mengetahui apakah soal tergolong sangat mudah, mudah, sedang, sukar, atau sangat sukar. Sementara itu, distraktor dianalisis untuk mengetahui apakah pilihan jawaban pengecoh berfungsi dengan baik.

Tahap terakhir adalah interpretasi hasil analisis. Hasil analisis dari aplikasi Anates ditafsirkan untuk mengetahui kualitas setiap butir soal (Akbar et al., 2017). Butir soal yang memiliki kualitas baik dapat dipertahankan, sedangkan butir soal yang kurang baik perlu direvisi atau diganti. Dengan demikian, penelitian ini bertujuan untuk memperoleh gambaran mengenai mutu instrumen evaluasi pembelajaran serta memberikan dasar bagi pendidik dalam menyusun soal yang lebih baik, objektif, dan sesuai dengan kemampuan peserta didik.

C.Hasil Penelitian dan Pembahasan

Analisis Kualitas Distraktor pada Butir Soal

No Butir	a	b	c	d	*
1	18---	2**	0--	0--	0
2	19**	1---	0--	0--	0
3	0--	14**	4--	2++	0
4	1**	2-	1--	14---	2
5	17---	0--	1**	2-	0
6	2+	0--	16**	2+	0
7	0--	19**	1---	0--	0
8	0--	17**	2--	1++	0
9	1---	0--	19**	0--	0
10	1+	18**	1+	0--	0
11	19**	1---	0--	0--	0
12	2+	16**	2+	0--	0
13	0	20**	0	0	0
14	2-	8+	6++	4**	0
15	10**	6--	2+	2+	0
16	14**	4--	0--	0--	2
17	0	20**	0	0	0
18	2+	0--	2+	16**	0
19	20**	0	0	0	0
20	4**	4+	12---	0--	0

Gambar 1. Distraktor

Hasil analisis kualitas pengecoh menunjukkan bahwa setiap butir soal memiliki efektivitas distraktor yang berbeda-beda. Analisis ini dilakukan terhadap 20 butir soal dengan jumlah subjek sebanyak 20 peserta didik. Dalam hasil Anates, tanda ** menunjukkan kunci jawaban, sedangkan kualitas pengecoh dikategorikan menjadi sangat baik, baik, kurang, buruk, dan sangat buruk (Rojen & Nivea, 2024).

Kualitas distraktor yang cukup baik tampak pada beberapa butir soal, seperti butir 6, 10, 12, 14, 15, dan 18.

Pada butir-butir tersebut terdapat pilihan jawaban pengecoh yang masuk kategori baik atau sangat baik. Hal ini menunjukkan bahwa beberapa pilihan jawaban salah mampu menarik perhatian peserta didik, sehingga distraktor tersebut dapat dikatakan berfungsi dalam mengukur pemahaman peserta didik.

Namun, masih terdapat beberapa butir soal yang memiliki kualitas distraktor rendah. Hal ini terlihat pada butir 1, 2, 4, 5, 7, 9, 11, 13, 16, 17, dan 19, yang sebagian besar pilihan pengecohnya berada pada kategori buruk atau sangat buruk. Pada butir 13, 17, dan 19, pilihan pengecoh tidak dipilih oleh peserta didik karena seluruh peserta didik cenderung memilih kunci jawaban. Kondisi tersebut menunjukkan bahwa soal kemungkinan terlalu mudah atau pilihan pengecohnya kurang berfungsi karena mudah dikenali sebagai jawaban salah (Girish et al., 2023) .

Selain itu, terdapat beberapa butir soal yang hanya memiliki satu pengecoh yang berfungsi baik, sedangkan pilihan lainnya masih kurang efektif. Keadaan ini menunjukkan bahwa kualitas pilihan jawaban belum merata. Distraktor yang baik seharusnya disusun secara logis, seimbang, dan memiliki kedekatan dengan jawaban benar agar mampu mengukur pemahaman peserta didik secara lebih tepat (Yooseop et al., 2025).

Dengan demikian, kualitas distraktor pada instrumen tes ini masih bervariasi. Beberapa distraktor sudah berfungsi dengan baik dan dapat dipertahankan, sedangkan pengecoh yang tidak dipilih atau berkategori buruk perlu direvisi. Perbaikan dapat dilakukan dengan menyusun pilihan jawaban yang lebih masuk akal, tidak terlalu mudah ditebak, dan sesuai dengan materi yang diujikan.

Analisis Validitas pada Butir Soal

No Butir	Korelasi	Signifikansi
1	0.006	-
2	0.438	Signifikan
3	0.600	Sangat Signifikan
4	0.205	-
5	0.205	-
6	0.736	Sangat Signifikan
7	0.438	Signifikan
8	0.410	-
9	0.438	Signifikan
10	0.403	-
11	-0.044	-
12	0.736	Sangat Signifikan
13	NAN	NAN
14	0.359	-
15	0.683	Sangat Signifikan
16	0.906	Sangat Signifikan
17	NAN	NAN
18	0.736	Sangat Signifikan
19	NAN	NAN
20	0.403	-

Gambar 2. Validitas

Hasil analisis validitas menunjukkan bahwa setiap butir soal

memiliki tingkat keterkaitan yang berbeda terhadap skor total. Validitas butir dalam penelitian ini dilihat dari nilai korelasi antara skor pada masing-masing butir dengan skor keseluruhan peserta didik. Semakin besar nilai korelasi yang dihasilkan, maka semakin baik pula kemampuan butir soal tersebut dalam mengukur kompetensi peserta didik sesuai tujuan evaluasi (Siti et al., 2023) .

Dari 20 butir soal yang dianalisis, terdapat sejumlah butir yang termasuk dalam kategori signifikan dan sangat signifikan. Butir soal nomor 2, 7, dan 9 memperoleh nilai korelasi 0,438 dan termasuk kategori signifikan. Sementara itu, butir soal nomor 3 memperoleh korelasi 0,600, nomor 6, 12, dan 18 memperoleh korelasi 0,736, nomor 15 memperoleh korelasi 0,683, serta nomor 16 memperoleh korelasi tertinggi, yaitu 0,906. Butir-butir tersebut menunjukkan bahwa soal memiliki hubungan yang baik dengan skor total, sehingga dapat dikatakan memiliki validitas yang cukup kuat.

Di sisi lain, masih ditemukan beberapa butir soal yang belum mencapai kategori signifikan, yaitu butir nomor 1, 4, 5, 8, 10, 11, 14, dan 20. Nilai korelasi pada butir-butir tersebut masih tergolong rendah. Bahkan, butir nomor 11 memiliki nilai korelasi negatif sebesar -0,044. Hal ini menunjukkan bahwa butir tersebut kurang mampu mengukur kemampuan peserta didik secara tepat dan perlu ditinjau kembali.

Selain itu, terdapat tiga butir soal yang nilai korelasinya tidak dapat dihitung atau menunjukkan hasil NAN, yaitu butir nomor 13, 17, dan 19. Keadaan ini kemungkinan terjadi karena jawaban peserta didik pada butir tersebut tidak bervariasi, misalnya hampir seluruh peserta didik memilih jawaban yang sama. Oleh sebab itu, butir dengan hasil NAN perlu diperhatikan secara khusus karena belum dapat memberikan informasi validitas yang jelas.

Secara umum, hasil analisis validitas memperlihatkan bahwa terdapat 9 butir soal yang memiliki validitas baik, terdiri atas 3 butir signifikan dan 6 butir sangat signifikan. Sementara itu, 8 butir soal belum signifikan dan 3 butir soal tidak dapat dihitung nilai korelasinya. Dengan demikian, butir soal yang signifikan dan sangat signifikan dapat dipertahankan, sedangkan butir soal yang belum signifikan maupun tidak dapat dihitung perlu direvisi agar

instrumen evaluasi menjadi lebih berkualitas (Agustina et al., 2023) .

Analisis Daya Pembeda pada Butir Soal

No Butir	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	1	1	0	0.00
2	5	4	1	20.00
3	4	1	3	60.00
4	1	0	1	20.00
5	1	0	1	20.00
6	5	2	3	60.00
7	5	4	1	20.00
8	5	3	2	40.00
9	5	4	1	20.00
10	5	4	1	20.00
11	5	5	0	0.00
12	5	1	4	80.00
13	5	5	0	0.00
14	1	0	1	20.00
15	4	0	4	80.00
16	5	0	5	100.00
17	5	5	0	0.00
18	5	1	4	80.00
19	5	5	0	0.00
20	3	0	3	60.00

Gambar 3. Daya Pembeda

Daya pembeda butir soal memperlihatkan kemampuan setiap soal dalam membedakan peserta didik yang memiliki kemampuan tinggi dan peserta didik yang memiliki kemampuan rendah (Rezigalla et al., 2024). Aspek ini penting dalam analisis butir soal karena dapat menunjukkan apakah suatu soal mampu mengukur perbedaan tingkat penguasaan peserta didik secara tepat (Rabiatul, 2021).

Analisis dilakukan pada 20 butir soal dengan jumlah subjek sebanyak 20 peserta didik. Dalam proses analisis, peserta didik dibagi menjadi kelompok atas dan kelompok bawah, masing-masing berjumlah 5 peserta didik. Indeks daya pembeda pada setiap butir soal menunjukkan variasi, mulai dari 0,00% sampai 100,00%.

Butir soal yang memiliki daya pembeda paling tinggi terdapat pada nomor 16 dengan indeks sebesar 100,00%. Selain itu, butir nomor 12, 15, dan 18 memperoleh indeks daya pembeda sebesar 80,00%. Nilai tersebut menunjukkan bahwa butir-butir soal tersebut mampu membedakan peserta didik yang menguasai materi dengan peserta didik yang belum menguasai materi. Oleh karena itu, butir soal tersebut dapat dikategorikan baik dan layak untuk dipertahankan.

Selanjutnya, butir nomor 3, 6, dan 20 memperoleh indeks daya pembeda sebesar 60,00%. Hal ini menunjukkan bahwa ketiga butir tersebut juga memiliki kemampuan yang baik dalam membedakan kelompok atas dan kelompok bawah. Jumlah peserta didik kelompok atas yang menjawab benar lebih banyak dibandingkan kelompok bawah, sehingga butir soal tersebut masih dapat digunakan dalam evaluasi pembelajaran.

Butir nomor 8 memiliki indeks daya pembeda sebesar 40,00%. Nilai ini menunjukkan bahwa soal tersebut masih cukup mampu membedakan

kemampuan peserta didik, meskipun kualitasnya perlu ditinjau kembali agar menjadi lebih baik. Sementara itu, butir nomor 2, 4, 5, 7, 9, 10, dan 14 memperoleh indeks daya pembeda sebesar 20,00%. Indeks tersebut menunjukkan bahwa kemampuan soal dalam membedakan peserta didik masih tergolong rendah, sehingga butir-butir tersebut perlu diperbaiki.

Terdapat beberapa butir soal yang memperoleh indeks daya pembeda 0,00%, yaitu butir nomor 1, 11, 13, 17, dan 19. Butir-butir tersebut belum mampu membedakan peserta didik kelompok atas dan kelompok bawah karena jumlah peserta didik yang menjawab benar pada kedua kelompok sama. Kondisi ini menunjukkan bahwa soal kurang efektif dalam mengukur perbedaan kemampuan peserta didik, sehingga perlu direvisi atau diganti.

Secara umum, daya pembeda pada instrumen tes ini menunjukkan bahwa sebagian butir soal telah memiliki kualitas yang baik, terutama butir dengan indeks daya pembeda tinggi. Namun, masih terdapat beberapa butir yang perlu diperbaiki karena memiliki daya pembeda rendah atau bahkan tidak memiliki daya pembeda. Perbaikan dapat dilakukan dengan menelaah kembali redaksi soal, kesesuaian materi, tingkat kesukaran, serta pilihan jawaban agar soal lebih tepat dalam mengukur kemampuan peserta didik.

Analisis Tingkat Kesukaran Butir Soal

No Butir	Jml Betul	Tkt. Kesukaran(%)	Tafsiran
1	2	10.00	Sangat Sukar
2	19	95.00	Sangat Mudah
3	14	70.00	Sedang
4	1	5.00	Sangat Sukar
5	1	5.00	Sangat Sukar
6	16	80.00	Mudah
7	19	95.00	Sangat Mudah
8	17	85.00	Mudah
9	19	95.00	Sangat Mudah
10	18	90.00	Sangat Mudah
11	19	95.00	Sangat Mudah
12	16	80.00	Mudah
13	20	100.00	Sangat Mudah
14	4	20.00	Sukar
15	10	50.00	Sedang
16	14	70.00	Sedang
17	20	100.00	Sangat Mudah
18	16	80.00	Mudah
19	20	100.00	Sangat Mudah
20	4	20.00	Sukar

Gambar 4. Tingkat Kesukaran

Tingkat kesukaran digunakan untuk melihat apakah soal yang diberikan termasuk mudah, sedang, atau sulit bagi peserta didik (Nguyen et al., 2023). Dari hasil yang diperoleh, setiap butir soal memiliki tingkat kesukaran yang berbeda-beda. Ada soal yang dapat dijawab dengan benar oleh banyak peserta didik, tetapi ada juga soal yang hanya dapat dijawab oleh sedikit peserta didik (Luthfiah et al., 2023).

Pada analisis ini terdapat 20 butir soal yang dikerjakan oleh 20 peserta

didik. Beberapa soal termasuk sangat sukar, yaitu soal nomor 1, 4, dan 5. Soal nomor 1 hanya dijawab benar oleh 2 peserta didik, sedangkan soal nomor 4 dan 5 hanya dijawab benar oleh 1 peserta didik. Hal ini menunjukkan bahwa soal-soal tersebut cukup sulit dikerjakan, sehingga perlu diperiksa kembali, baik dari segi bahasa soal, materi, maupun pilihan jawabannya.

Soal yang termasuk kategori sukar terdapat pada nomor 14 dan 20. Kedua soal tersebut hanya dijawab benar oleh 4 peserta didik. Soal seperti ini masih bisa digunakan, tetapi perlu diperhatikan kembali agar tidak terlalu menyulitkan peserta didik. Apabila soal dibuat untuk mengukur pemahaman dasar, maka soal tersebut sebaiknya diperbaiki agar lebih mudah dipahami.

Selanjutnya, soal nomor 3, 15, dan 16 termasuk dalam kategori sedang. Soal nomor 3 dan 16 dijawab benar oleh 14 peserta didik, sedangkan soal nomor 15 dijawab benar oleh 10 peserta didik. Soal dengan kategori sedang dapat dikatakan cukup baik karena tidak terlalu mudah dan tidak terlalu sulit. Soal seperti ini lebih seimbang digunakan dalam evaluasi.

Soal yang termasuk kategori mudah terdapat pada nomor 6, 8, 12, dan 18. Soal-soal tersebut dapat dijawab benar oleh sebagian besar peserta didik. Hal ini menunjukkan bahwa peserta didik cukup mampu memahami materi yang ditanyakan pada soal tersebut. Meskipun termasuk mudah, soal-soal ini masih dapat digunakan, terutama untuk mengukur pemahaman dasar peserta didik.

Sementara itu, soal nomor 2, 7, 9, 10, 11, 13, 17, dan 19 termasuk kategori sangat mudah. Bahkan soal nomor 13, 17, dan 19 dijawab benar oleh seluruh peserta didik. Soal yang terlalu mudah kurang mampu membedakan kemampuan peserta didik, karena hampir semua peserta didik dapat menjawabnya dengan benar. Oleh karena itu, soal-soal tersebut perlu ditinjau kembali agar kualitasnya lebih baik.

Secara keseluruhan, tingkat kesukaran soal dalam tes ini cukup beragam. Ada soal yang terlalu sulit, ada yang sedang, dan ada pula yang terlalu mudah. Soal dengan tingkat kesukaran sedang dapat dipertahankan, sedangkan soal yang sangat sukar dan sangat mudah sebaiknya diperbaiki (Ika et al., 2021).

Perbaikan ini penting agar soal yang digunakan benar-benar sesuai dengan kemampuan peserta didik dan tujuan evaluasi pembelajaran.

E. Kesimpulan

Analisis butir soal berbasis aplikasi Anates sangat membantu dalam mengetahui kualitas instrumen tes yang digunakan dalam evaluasi pembelajaran. Instrumen tes yang dianalisis memiliki reliabilitas tinggi, yaitu sebesar 0,87, sehingga secara umum tes tersebut memiliki tingkat konsistensi yang baik.

Namun, kualitas setiap butir soal masih bervariasi. Dari segi validitas, terdapat beberapa butir soal yang sudah layak dipertahankan karena memiliki hubungan yang baik dengan skor total, tetapi masih ada beberapa butir yang belum signifikan dan perlu direvisi. Dari segi daya pembeda, sebagian soal mampu membedakan peserta didik berkemampuan tinggi dan rendah, sedangkan beberapa soal belum memiliki daya pembeda yang baik. Tingkat kesukaran soal juga menunjukkan variasi, mulai dari sangat sukar, sedang, mudah, hingga sangat mudah. Selain itu, kualitas distraktor pada beberapa soal sudah berfungsi dengan baik, tetapi masih terdapat pengecoh yang kurang efektif karena tidak dipilih oleh peserta didik.

Dengan demikian, butir soal yang memiliki validitas baik, daya pembeda tinggi, tingkat kesukaran sedang, dan distraktor yang berfungsi dapat dipertahankan. Sementara itu, butir soal yang terlalu mudah, terlalu sukar, memiliki daya pembeda rendah, atau distraktornya tidak berfungsi perlu direvisi atau diganti.

Hasil analisis ini dapat dijadikan bahan evaluasi bagi pendidik agar penyusunan soal pada pembelajaran berikutnya menjadi lebih baik, objektif, dan sesuai dengan tujuan evaluasi pembelajaran.

DAFTAR PUSTAKA

Adom, D., Adu-Mensah, J., & Dake, D. A. (2020). Tes, pengukuran, dan evaluasi: Pemahaman dan penggunaan konsep dalam pendidikan. *International Journal of Evaluation and Research in Education*. https://api.semanticscholars.org/CorpusID:216286878

Agustina, S. T. S., Rosane, M., & Eko, R. (2023). Analisis Kualitas dan Tingkat Kognitif Soal Penilaian Akhir Fisika Kelas X. *IJOEM Indonesian Journal of E-Learning and Multimedia Multimedia*.

Akbar, M., Firman, H., & Rusyati, L. (2017). Mengembangkan Tes Virtual Sains untuk Mengukur

Pemikiran Kritis Siswa tentang Makhluk Hidup dan Tema Keberlanjutan Lingkungan. *Journal of Physics: Conference Series*, *812*. https://api.semanticscholars.org/CorpusID:151540483

Arianda, Y. D., Triana, D. D., & Madani, F. (2024). Mengatasi Kesenjangan dalam Evaluasi Pendidikan: Sebuah Meta-Analisis Komprehensif. *Mimbar Ilmu*. https://api.semanticscholars.org/CorpusID:277155868

Bulala, S. N. dan S. A. dan T. (2023). Estimasi Sifat Psikometrik (Kesulitan Butir Soal, Diskriminasi, dan Indeks Reliabilitas. *Shanlax International Journal of Education*. https://api.semanticscholar.org/CorpusID:260180563

Girish, C. R., Bhoomika, C. R., Jayesh, V. V., & Pradip, C. R. (2023). Hubungan Jumlah Distraktor yang Berfungsi dengan Indeks Kesulitan Butir Soal dan Daya Diskriminasi Butir Soal dalam Soal Pilihan Ganda. *Cureus*.

Ika, M., Zuhdan, P. K., Insih, W., Siwi, P., & Meita, F. (2021). Soal Pilihan Ganda dan Esai HOTs: Instrumen Tervalidasi untuk Mengukur Keterampilan Berpikir Tingkat Tinggi Calon Guru. *Jurnal Pendidikan Sains Turki*.

Lorika, R., Arsih, F., Alberida, H., & Irdawati. (2024). *Kualitas Pertanyaan Argumentasi pada Materi Sistem Ekskresi Menggunakan Aplikasi ANATES*. https://api.semanticscholars.org/CorpusID:276653768

Luthfiah, Y., A, N., Nanda, A., Fhahnul, A. A. H., & Fadjryani. (2023). MENGANALISIS KUALITAS INSTRUMEN PENGUKURAN SOAL PILIHAN GANDA PADA MATERI EKONOMI KELAS XI DI SEKOLAH MENENGAH ATAS NEGERI 3 GORONTALO MELALUI TEORI UJI KLASIK DAN MODEL RASCH. *Parameter: Jurnal Statistik*.

Nabilah, A. M., Santika, M., & Nadir, N. (2025). Penerapan anates dalam menilai keberhasilan pembelajaran materi idgham dan iqlab di madrasah ibtidaiyah. *Mitra PGMI: Jurnal Kependidikan MI*. https://api.semanticscholar.org/CorpusID:277034050

Nguyen, S. M., Anh, H., Stamper, T. C., & C., J. (2023). *Menilai Kualitas Soal Pilihan Ganda Menggunakan GPT-4 dan Metode Berbasis Aturan*. https://api.semanticscholars.org/CorpusID:259937876

Nur, F., & Hartono. (2023). Analisis Kelayakan Butir Soal Penilaian Akhir Semester Mata Pelajaran Bahasa Indonesia Kelas VII SMPN 1 Ngaglik. *Britain International of Linguistics Arts and Education (BIoLAE) Journal*.

Rabiatul, A. (2021). ANALISIS BERBASIS APLIKASI TERHADAP KUALITAS BUTIR SOAL DI SEKOLAH MENENGAH PERTAMA. *JISAE : Jurnal Penilaian Dan Evaluasi Siswa Indonesia*.

Rezigalla, A., AM, E., A., E., A., A., M, A., Ahmed, A. A. Y., A, Y., Osama, M. A., & MI, A. (2024).

Analisis butir soal: dampak efisiensi distraktor terhadap indeks kesulitan dan daya diskriminasi butir soal pilihan ganda. *BMC Medical Education*.

Rojen, D. M. N., & Nivea, S. L. D. (2024). Analisis Efisiensi Pilihan Ganda dalam Ujian Komprehensif Pilihan Ganda untuk Calon Guru Pendidikan Kejuruan. *Jurnal Internasional Untuk Penelitian Multidisiplin*.

Siti, Q., Rani, D., Ummi, R., & Irma, A. (2023). Indikator dan Kisi-kisi Soal Esai Materi Tiga Dimensi: Pengembangan Instrumen Pengukuran Kemampuan Pemecahan Masalah Matematis Siswa SMA. *JEMS: Jurnal Edukasi Matematika Dan Sains*.

Sunismi, M. M., Fuady, A., & Sunismi. (2023). *Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum Kabupaten Malang*. https://api.semanticscholar.org/CorpusID:259821771

Yooseop, L., Suin, K., & Yohan, J. (2025). *Menghasilkan Pilihan Jawaban yang Masuk Akal untuk Soal Pilihan Ganda melalui Prediksi Pilihan Siswa*.

LAMPIRAN











