GRAVITASI
Jurnal Pendidikan Fisika dan Sains
Vol (8) No (02) Edisi Desember Tahun 2025

https://ejurnalunsam.id/index.php/JPFS

Analisis Kualitas Instrumen Tes Fisika Kelas X Dengan Memanfaatkan Aplikasi AnBuso

Wahyu Putri Mustikasari[1], Raden Wakhid Akhdinirwanto [1]

[1]Program Studi Pendidikan Fisika FKIP Universitas Muhammadiyah Purworejo
Jln. K.H. Ahmad Dahlan No. 3, Kecamatan Purworejo, Kabupaten Purworejo, Provinsi Jawa Tengah, 54111
Email Korespondensi: aissspuput@gmail.com

ABSTRAK

Penelitian bertujuan untuk mengkaji instrumen tes Fisika kelas X dengan memanfaatkan apikasi AnBuso sebagai alat bantu analisis. Instrumen tes yang digunakan adalah ulangan akhir semester dengan 15 soal pilihan ganda. Proses analisis dilakukan dari segi validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Temuan penelitian menunjukkan bahwa 53,33% dari soal yang dianggap valid, sedangkan 7 soal lainnya tidak valid, dan instrumen tes menunjukkan memiliki kemampuan reliabilitas sebesar 0,78 (tinggi). Sebanyak 93,33% soal tergolong mudah, 6,67% sedang, dan tidak ada soal yang tergolong sulit. Daya pembeda menunjukkan bahwa 66,67% soal memiliki fungsi pembeda yang baik, 13,33% cukup baik, dan 20% tidak baik. Dengan demikian, instrumen tes yang digunakan memiliki validitas yang cukup memadai dan reliabilitas yang tinggi, namun cenderung tergolong kategori soal yang mudah serta masih memerlukan perbaikan di aspek daya pembeda. Soal-soal yang tidak memenuhi standar validitas serta memiliki kemampuan daya pembeda yang rendah perlu diperbaiki atau diganti supaya tes lebih efektif dalam menilai kemampuan siswa dengan cara yang objektif dan tepat. Hasil analisis ini diharapkan dapat menjadi pedoman untuk para guru dalam meningkatkan kualitas butir soal evaluasi di kelas.

Kata kunci: Analisis Kualitas Soal, Aplikasi Anbuso, Instrumen Tes Fisika

ABSTRACT

The study aims to examine the Physics test instrument for grade X by utilizing the AnBuso application as an analysis tool. The test instrument used was a final semester exam with 15 multiple-choice questions. The analysis process was carried out in terms of validity, reliability, difficulty level, and discriminatory power. The research findings showed that 53.33% of the questions were considered valid, while 7 other questions were invalid, and the test instrument showed a reliability capability of 0.78 (high). A total of 93.33% of the questions were classified as easy, 6.67% were moderate, and no questions were classified as difficult. The discriminatory power showed that 66.67% of the questions had a good discriminatory function, 13.33% were quite good, and 20% were not good. Thus, the test instrument used has sufficient validity and high reliability, but tends to be classified as an easy question category and still requires improvement in the aspect of discriminatory power. Questions that do not meet validity standards and have low discriminatory power need to be revised or replaced to make the test more effective in assessing student abilities objectively and accurately. The results of this analysis are expected to serve as a guide for teachers in improving the quality of evaluation items in the classroom.

Keywords: Question Quality Analysis, Anbuso Application, Physics Test Instrument

A. PENDAHULUAN

Mutu pendidikan adalah aspek yang sangat krusial dalam bidang pendidikan. Salah satu faktor utama yang menentukan kualitas pendidikan adalah evaluasi. Evaluasi adalah proses penilaian yang dilaksanakan dengan tujuan untuk mengetahui seberapa baik siswa mencapai target pembelajaran. (Maulidah et al., 2023). Evaluasi di dalam pendidikan

adalah elemen penting bagi proses belajar mengajar. Ketika pembelajaran dipahami sebagai suatu proses perubahan perilaku, maka fungsi evaluasi bisa dianggap sebagai sarana untuk mengetahui hal tersebut. Sistem evaluasi yang efektif akan memberikan gambaran mengenai kualitas proses belajar, yang pada akhirnya dapat membantu para pendidik dalam menyusun rencana pembelajaran (Fathoni, 2023). Tanpa ada penilaian, guru akan kesulitan dalam menilai kemajuan siswa. Dengan demikian, penilaian yang baik menjadi dasar penting dalam meningkatkan kualitas prosen belajar di sekolah.

Instrumen penilaian adalah alat yang dipakai untuk menilai pencapaian kompetensi, kapasitas, atau kualitas hasil belajar peserta didik. Instrumen penelitian yang efektif harus mencapai dua syarat penting, yaitu validitas dan reliabilitas. Validitas berhubungan dengan sejauh mana alat ukur dapat menilai sesuai dengan yang seharusnya dinilai, sedangkan reliabilitas berbicara tentang kestabilan hasil pengukuran jika dilakukan berulang kali (Purba et al., 2021). Instrumen yang valid mencerminkan kondisi nyata dari variabel yang diteliti, sementara instrumen yang reliabel memberikan data yang stabil dan konsisten dalam situasi yang sama (Zayrin et al, 2025). Oleh karena itu, alat evaluasi yang diterapkan perlu mengikuti ketentuan yang ada dan harus mampu menilai semua elemen kemampuan siswa. Dengan begitu, dibutuhkan alat evaluasi yang valid dan reli abel untuk mengukur sejauh mana efisiensi proses pembelajaran.

Dalam rangka menilai kualitas suatu ujian, analisis terhadap setiap soal harus dilakukan untuk menilai seberapa baik tes tersebut dapat berfungsi sebagai alat pengu kuran. Proses analisis ini dapat ditunjang oleh teknologi, salah satunya dengan pema nfaatan perangkat lunak AnBuso (Analisis Butir Soal). AnBuso merupakan program berbasis Microsoft Excel yang berfungsi untuk membantu guru dalam menganalisis hasil ujian siswa secara cepat dan praktis. Aplikasi AnBuSo (Analisis Butir Soal) telah dirancang sedemikian rupa sehingga guru tidak menemui kesulitan dalam menganalis is butir soal. Program ini sangat ramah pengguna dan dapat diterapkan oleh setiap guru karena pada dasarnya software ini menggunakan aplikasi pengolah angka yang sangat familiar bagi guru yaitu Microsoft Excel (Farhan et al, 2021). Program ini juga dapat diakses secara gratis (Adawiah et al., 2025). Dalam aplikasi ini juga tersedia untuk tes esai, pilihan ganda, serta kelompok perbaikan (Heriyati et al, 2023). Melalui AnBuso, guru dapat mengevaluasi validitas soal, tingkat kesukaran, daya pembeda dan reliabilitas dari setiap pertanyaan.

Meskipun analisis kualitas dari setiap pertanyaan merupakan tahap yang penting untuk memastikan hasil evaluasi yang akur at rupanya guru guru umumnya jarang me manfaatkan analisis kualitas butir soal secara optimal guna memanfaatkan kinerja yang lebih sempurna lagi. Banyak guru yang memang melangsungkan penilaian yang manual melibatkan waktu yang cukup lama serta rentan terjadinya kesalahan perhitungan yang berlebihan. Kondisi ini mengakibatkan hasil penilaian terhadap kemampuan siswa tidak mencerminkan keadaan yang sebenarnya (Adawiah et al., 2025). Di MAN Purworejo, banyak guru masih mengalami masalah saat ingin melakukan evaluasi yang efektif. Proses penilaian ini sering kali dilakukan dengan cara manual yang memakan waktu dan bisa membuat kesalahan. Sebagai hasilnya, nilai yang didapat tidak selalu menunjukkan kemampuan asli siswa, sehingga sulit untuk digunakan sebagai patokan dalam menentukan langkah pembelajaran yang benar. Oleh sebab itu, studi ini menjadi sangat penting untuk dilak sanakan guna mendapatkan pemahaman empiris tentang kelayakan instrumen tes fisika menggunakan software AnBuso dan membantu meningkatkan kualitas penilaian pembelajaran.

Tujuan penelitian ini adalah untuk mengevaluasi kualitas alat ukur tes Fisika bagi siswa kelas X dengan menggunakan software AnBuso untuk alat bantu analisisnya. Dalam melakukan riset aplikasi ini, maka pembuktian akan validitas, reliabilitas dan analisis tingkat kesulitan, daya pembeda butir soal diadakan untuk melakukan analisis ini dengan maksud agar hasil analisis ini berkontribusi untuk mengenali kelayakan

tes yang digunakan sehingga sebagai acuan dalam rangkaian kegiatan. Penelitian ini diharapkan memberi keuntungan bagi banyak pihak. Bagi para guru penelitian ini adalah dapat digunakan sebagai acuan atau pedoman dalam meningkatkan kualitas butir-butir soal yang diterapkan dalam proses evaluasi pembelajaran agar penilaian yang dilaksanakan lebih akurat dan tidak berdasarkan secara tidak teliti atau tanpa distorsi bahwa penilaian sebenarnya sudah tidak lagi sesuai lagi dengan kualitas atau kadar kemampuan siswa setelah dilakukan proses pembelajaran untuk keduanya, siswa dan guru, bahwa penilaian sudah melihat setiap siswa seperti individu yang berbeda-beda harus tahu dan harus paham penilaian.

B. METODE PENELITIAN

Metode penelitian ini menggunakan deskriptif kuantitatif dengan fokus utama pada siswa kelas X di MAN Purworejo pada materi Sistem Pengukuran dalam Kerja Ilmiah**.** Data penelitian berupa hasil tes siswa yang berjumlah 15 butir soal pilihan ganda. Penelitian dilaksanakan di MAN Purworejo pada 15 Juli 2025 sampai 4 Oktober 2025. Subjek penelitian adalah soal ulangan akhir Fisika kelas X sebanyak 15 butir soal, sementara itu objek penelitian ini adalah kualitas dari instrumen tes Fisika yang diterapkan. Data dikumpulkan melalui telaah dokumen, meliputi lembar soal dan jawaban siswa. Analisis kuantitatif dapat dilihat melalui aspek-aspek seperti validitas, tingkat kesukaran, daya pembeda, dan reliabilitas. Data yang diperoleh akan dianalisis menggunakan aplikasi AnBuso.

Validitas

Validitas menggambarkan bahwa sebuah tes dapat mengukur secara tepat apa yang seharusnya diukur. Uji validitas memiliki tujuan untuk mengevaluasi apakah alat ukur tersebut akurat atau tidak (Saputri et al, 2023). Salah satu metode untuk menilai validitas soal adalah melalui penggunaan korelasi skor validitas. Dalam evaluasi soal, AnBso memanfaatkan koefisien korelasi point biserial (rpbis). Perhitungan dilakukan dengan rumus di bawah ini:

$$r_{pbis} = \frac{(X_1 - X)\sqrt{p_1 - (1 - p_1)}}{s_X}$$

Keterangan:

X_1= Rata-rata nilai total ujian dari siswa yang menjawab dengan tepat

X = Rata-rata nilai total ujian dari semua siswa

p_1 = Persentase siswa yang menjawab pertanyaan dengan tepat

s_X = Deviasi standar nilai total ujian secara keseluruhan

Tabel 1. Kategori Validitas

Koefisien Validitas	Kategori
0,80-1,00	Sangat Tinggi
0,60-0,80	Tinggi
0,40-0,60	Cukup
0,20-0,40	Rendah
0,00-0,20	Sangat Rendah

(Sumber: Arikunto, 2015)

Reliabilitas

Reliabilitas berasal dari kata *reliability* yang mengacu pada hasil pengukuran bisa dipercaya. Suatu pengukuran dianggap dapat dipercaya jika, pada berbagai kesempatan menilai kelompok yang sama, hasil yang diperoleh cenderung stabil, asalkan variabel yang diukur pada subjek tersebut tetap tidak berubah (Farida, 2021). Koefisien reliabilitas dapat dihitung dengan metode Kuder-Richardson yang dikenal sebagai KR-20, melalui persamaan berikut:

$$\text{KR} - 20 = \frac{K}{K-1}\left(1 - \frac{\sum P_i q_i}{S_{i^2}}\right)$$

Keterangan:

K = Jumlah item

p_{iq_i} = Varian skor butir

p_i= Proporsi jawaban benar untuk soal nomor i

q_i= Proporsi jawaban salah untuk soal nomor

S^2= Total varians skor responden

Tabel 2. Kriteria Reliabilitas Soal

Koefisien Reliabilitas	Keterangan
0,00-0,20	Sangat Rendah
0,20-0,40	Rendah
0,40-0,70	Cukup
0,70-0,90	Tinggi
0,90-1,00	Sangat Tinggi

(Sumber: Arikunto, 2010)

Tingkat Kesukaran

Tingkat kesukaran merujuk kepada kemampuan dalam memilih atau mengatur pertanyaan yang kesulitannya selaras dengan kemampuan atau pengetahuan individu atau kelompok yang dihadapi

(Rajagukguk, 2023). Tingkat sulitnya soal bisa dihitung dengan rumus sebagai berikut:

$$TK = \frac{B}{N} x\ 100\%$$

Keterangan:
TK = Tingkat kesukaran
B = Jumlah siswa yang menjawab dengan benar
N = Total seluruh siswa

Tabel 3. Kriteria Tingkat Kesukaran

Kriteria Tingkat Kesukaran	Keterangan
0,00-0,30	Sukar
0,31-0,70	Sedang
0,70-1,00	Mudah

(Sumber: Arikunto, 2010)

Daya Pembeda

Daya pembeda adalah potensi suatu soal yang dapat mengklasifikasikan peserta ujian ke dalam kelompok atas dan kelompok bawah (Pinaka, 2024). Secara sederhana, semakin tinggi daya pembeda suatu soal, semakin banyak peserta dari kelompok atas yang dapat menjawab soal dimana semakin sedikit peserta dari kelompok yang lebih bawah yang mampu memberikan jawaban yang benar (Purba et al., 2021). Daya pembeda suatu soal dapat ditentukan melalui rumus berikut:

$$D = \frac{B_A - B_B}{N_A} x\ 100\%$$

Keterangan:
D = Daya pembeda
B_A= Total siswa dalam grub atas dengan jawaban benar
B_B= Total siswa dalam grub bawah dengan jawaban benar
N = Total siswa di salah satu grub A atau B

Tabel 4. Kategori Daya Pembeda

Koefisien Daya Pembeda	Kategori
0,40-1,00	Baik Sekali
0,30-0,39	Baik
0,20-0,29	Cukup Baik
0,00-0,19	Tidak Baik

(Sumber: Arikunto, 2010)

C. HASIL DAN PEMBAHASAN

Hasil

Berdasarkan temuan dari studi yang dilakukan di MAN Purworejo Kelas X, didapati 38 siswa dengan total 15 soal pilihan ganda, di mana setiap jawaban yang benar memperoleh nilai 1 dan dan setiap jawaban yang salah memperoleh nilai 0. Pengelompokan soal dilakukan dengan mengacu pada hasil analisis kuantitatif yang dihasilkan dari ulangan akhir untuk mata pelajaran Fisika di kelas X di MAN Purworejo. Skor penilaian Aplikasi AnBuso terlihat pada tabel 5.

Tabel 5. Data Hasil Ulangan Akhir Berdasarkan Skor Nilai

Jumlah Jawaban Benar	Jumlah Jawaban Salah	Jumlah Siswa	Persentase Jawaban Benar (%)
15	0	11	100%
14	1	5	93%
13	2	5	87%
12	3	6	80%
11	4	3	73%
10	5	1	67%
9	6	4	60%
8	7	1	53%
6	9	1	40%

Penilaian kualitas soal dilakukan dengan memperhatikan berbagai faktor, yaitu validitas item, daya diskriminasi, dan tingkat kesulitan. Hasil penilaian kualitas soal dari ulangan akhir untuk bidang studi Fisika Kelas X dalam Tahun Ajaran 2025/2026 di MAN Purworejo adalah sebagai berikut.

Validitas

Untuk menghitung validitas soal perlu adanya perangkat lunak berupa aplikasi AnBuso. Populasi penelitian adalah 38 siswa Kelas X MAN Purworejo. Dengan jumlah subjek (*N*) 38 dan menggunakan taraf signifikan (α) 5% (0,05), diperoleh nilai koefisien r_{tabel} sebesar sebesar 0,320 (dengan derajat kebebasan dk = N-2 = 36). Kriteria keputusan validitas yang digunakan adalah soal yang dianggap valid jika $r_{hitung} \geq r_{tabel}$ (yaitu $r_{hitung} \geq 0{,}320$) dan dinyatakan tidak valid jika $r_{hitung} < r_{tabel}$ (yaitu $r_{hitung} < 0{,}320$).
Berdasarkan perhitungan dari total 15 butir soal, didapatkan rincian sebagai berikut:

Tabel 6. Hasil Perhitungan Validitas

Keterangan	Jumlah Butir Soal	Persentase (%)	Nomor Butir Soal
Valid	8	53,33%	3, 4, 6, 10, 11, 13, 14, 15
Tidak Valid	7	46,67%	1, 2, 5, 7, 8, 9, 12
Total	**15**	**100%**	

Tabel 7. Hasil Uji Validitas

Koefisien Validitas	Kategori	Jumlah	Item	Persentase (%)
0,80-1,00	Sangat Tinggi	0	-	0%
0,60-0,80	Tinggi	1	14	6,67%
0,40-0,60	Cukup	5	3, 6, 10, 11, 13	33,3%
0,20-0,40	Rendah	6	1, 4, 7, 15	40,00%
0,00-0,20	Sangat Rendah	3	2, 5, 8, 9, 12	20,00%
Total		**15**		**100%**

Reliabilitas

Dalam penelitian ini, reliabilitas instrumen tes fisika dianalisis menggunakan aplikasi AnBuso. Dari perhitungan yang dilakukan, didapatkan nilai koefisien reliabilitas mencapai angka 0,78. Nilai ini menandakan bahwa instrumen tes Fisika yang dipergunakan mempunyai tingkat reliabilitas yang bagus, karena nilai koefisien ini melebihi angka 0,70 (kriteria minimal yang umum digunakan untuk instrumen yang reliabel. Nilai koefisien reliabilitas 0,78 termasuk tinggi (>0,70) berdasarkan rumus KR-20 via AnBuso, menandakan konsistensi pengukuran.

Tingkat Kesukaran

Dalam analisis tingkat kesukaran, koefisien (P) menunjukkan persentase siswa yang memberikan jawaban benar. Tingkat kesukaran dari suatu soal dapat diidentifikasi dari berapa banyak siswa yang memberikan jawaban dengan tepat, siswa mampu menjawab pertanyaan tersebut. Apabila semakin banyak siswa yang keliru dalam menjawab, maka soal itu akan dikategorikan sebagai sulit atau rumit (Surya et al, 2023).

Tabel 8. Hasil Uji Akhir Berdasarkan Tingkat Kesukaran

Indeks Kesukaran	Kategori	Jumlah (Butir Soal)	Nomor Butir Soal	Persentase (%)
$0,00 < P \leq 0,30$	Sukar	0	-	0%
$0,31 < P \leq 0,70$	Sedang	1	15	6,67%
$0,71 < P \leq 1,00$	Mudah	14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	93,3%
Total		**15**		**100%**

Daya Pembeda

Sebuah instrumen tes juga harus melalui pengujian kemampuan membedakannya. Pengujian kemampuan membedakan ini adalah seberapa efektif suatu alam harus mampu membedakan siswa dengan kemampuaan tinggi dengan mereka yang memiliki kemampuan rendah (Surya et al, 2023).

Tabel 9. Hasil Uji Daya Pembeda

Koefisien Daya Beda	Kategori	Jumlah Soal	Nomor Butir Soal	Persentase %
0,40-1,00	Baik Sekali	3	3, 6, 10, 11 13, 14	20,00%
0,30-0,39	Baik	7	4, 7, 15	46,67%
0,20-0,29	Cukup Baik	2	1, 2, 8	13,33%
0,00-0,19	Tidak Baik	3	5, 9, 2	20,00%
Total		**15**		**100%**

Pembahasan

Validitas

Hasil analisis menunjukkan bahwa instrumen tes Ulangan Akhir Fisika Kelas X MAN Purworejo secara keseluruhan memiliki validitas yang dapat diterima karena 53,33% dari butir soal dinyatakan valid. Sejalan dengan temuan Yolanda (2023) yang menganalisis 20 butir soal ulangan harian Fisika dan menemukan hanya 30% butir valid (r-hitung 0,61-0,65),

menunjukkan bahwa validitas instrumen tes Fisika SMA sering kali perlu revisi signifikan pada butir dengan korelasi rendah. Butir soal yang valid tersebut tersebar pada kategori Tinggi, Cukup, dan sebagian kategori Rendah (karena nilainya masih ≥ 0,320).

Sebanyak 7 butir soal (46,67%) dinyatakan tidak dianggap valid. Soal yang tidak dianggap valid ini didominasi oleh soal yang tergolong dalam kategori rendah dan sangat rendah (termasuk butir 5, 9, dan 12 yang memiliki r_{hitung} sangat kecil). Nilai r_{hitung}yang rendah mengindikasikan bahwa kemampuan siswa dalam menjawab soal-soal tersebut tidak konsisten dengan skor total yang diperoleh oleh para siswa. Dengan kata lain, soal-soal ini tidak berfungsi dengan baik dalam memisahkan siswa dengan kemampuan tinggi dari mereka yang memiliki kemampuan lebih rendah.

Tingkat validitas yang sangat rendah dari butir soal dapat disebabkan oleh kurangnya pembuatan kisi-kisi soal oleh guru sebelum menyusun soal ulangan harian, sehingga soal yang dibuat bersifat asal-asalan dan tidak mencakup materi yang diajarkan. Soal yang tidak terpercaya akan memengaruhi hasil yang diperoleh oleh siswa (Yolanda, 2023).

Berdasarkan temuan ini, butir soal berangka 1, 2, 5, 7, 8, 9, dan 12 perlu ditinjau ulang, direvisi, atau dibuang sebelum digunakan kembali. Butir soal dengan validitas sangat rendah (terutama nomor 5, 9, dan 12) disarankan untuk diganti karena tidak berfungsi dengan baik sebagai pengukur pengetahuan siswa.

Reliabilitas

Hasil ini konsisten dan stabil untuk pengujian ulang, sesuai kriteria (Zayrin et al, 2025) yang menetapkan ambang reliabel ≥0,70 pada instrumen fisika. Reliabilitas tinggi memungkinkan pengukuran objektif kemampuan siswa.

Nilai reliabilitas yang mencapai 0,78 menunjukkan bahwa instrumen tes fisika untuk kelas X di MAN Purworejo dapat menghasilkan data yang konsisten dan stabil saat diterapkan berulang kali dalam kondisi yang serupa. Dengan kata lain, ketika siswa menjalani ujian dalam situasi yang serupa pada waktu yang berbeda, hasil yang diperoleh kemungkinan besar tidak akan menunjukkan perbedaan signifikan. Ini menunjukkan bahwa alat uji ini dapat diandalkan untuk menilai kemampuan siswa secara objektif dan konsisten.

Nilai reliabilitas tes Ulangan Akhir Fisika Kelas X MAN Purworejo mencapai 0,78 (kategori Tinggi), memenuhi kriteria Zayrin (2025) yang menetapkan batas keandalan ≥ 0,70 untuk alat ukur Fisika SMA dengan menggunakan Cronbach's Alpha. Nilai ini menunjukkan adanya konsistensi internal yang solid, dimana alat ukur menghasilkan skor yang stabil saat diterapkan berulang kali pada populasi yang serupa, dengan varians kesalahan pengukuran yang minimal.

Tingkat reliabilitas yang tinggi ini sejalan dengan hasil penelitian Hartono et al. (2024) yang menemukan r = 0,75 pada evaluasi Fisika SMA. Implikasinya, instrumen ini dapat dipercaya untuk penilaian Kurikulum Merdeka, dimana konsistensi pengukuran menjadi penting untuk membentuk profil pelajar Pancasila.

Tingkat Kesukaran

Hasil analisis tingkat kesulitan (P) dari instrumen tes Ulangan Akhir Fisika Kelas X MAN Purworejo menunjukkan kecenderungan yang signifikan ke arah kemudahan ekstrem, di mana 93,33% butir (14/15 soal) termasuk ke dalam kategori Mudah (P > 0,70).

1. Dominasi Soal Mudah: Sebanyak 93,33% (14 dari 15 butir soal) tergolong Mudah (P > 0,70). Dominasi soal mudah (P>0,70)) pada tes Fisika yang 85% butir tergolong mudah, menyebabkan instrumen kurang menantang siswa berprestasi tinggi (Hartono et al., 2024). Artinya, lebih dari 70% siswa dapat dengan akurat menjawab soal-soal ini.
2. Keseimbangan yang Rendah: Hanya 1 butir soal (6,67%) yang masuk kategori Sedang (butir 15 dengan P = 0,605). Butir tersebut adalah satu-satunya yang seimbang, sementara tidak ada soal yang sulit (P ≤ 0,30) menyebabkan tes ini kurang mampu mendeteksi kesulitan belajar siswa yang sebenarnya.

3. Absennya Soal Sukar: Tidak terdapat soal yang termasuk dalam kategori sulit (P ≤ 0,30). Berdasarkan analisis tingkat kesukaran dapat diambil kesimpulan bahwa tidak ada soal yang termasuk dalam klasifikasi sulit. Seluruh soal memiliki tingkat kesukaran lebih dari 0,30. Sehingga soal ini tidak mengandung soal dengan tingkat kesulitan yang tinggi.

Daya Pembeda

Hasil dari analisis daya pembeda menunjukkan bahwa secara keseluruhan, instrumen pengujian yang digunakan memiliki kapasitas yang cukup baik dalam membedakan level kemampuan akademik para siswa. Sebagian besar soal mampu membedakan antara siswa dengan kemampuan tinggi dan mereka yang memiliki kinerja rendah, meskipun masih ada beberapa soal yang perlu diperbaiki dan dihilangkan.

1. Soal yang Efektif (66,67%): Sebanyak 10 butir soal (gabungan kategori Baik Sekali dan Baik) memiliki fungsi pembeda yang baik dan layak dipertahankan. Butir-butir ini sukses membuat siswa yang memiliki skor total tinggi menjawab benar, sementara siswa dengan skor total rendah menjawab salah. Kondisi ini menunjukkan bahwa soal-soal tersebut telah memnuhi karakteristik butir tes yang berkualitas dan layak dipertahankan serta digunakan kembali dalam evaluasi pemelajaran.
2. Soal yang Marginal (13,33%): Butir soal kategori Cukup Baik (nomor 1, 2, dan 8) perlu direvisi. Meskipun masih berfungsi untuk membedakan, efektivitasnya dalam memishkan siswa dengan kemampuan tinggi dari yang rendah belum mencapai hasil yang diharapkan. Oleh sebab itu, disarankan untuk melakukan revisi pada soal-soal terutama pada pilihan pengecoh (distraktor) agar lebih berfungsi dan menarik perhatian siswa kelompok bawah, tetapi mudah diidentifikasi siswa kelompok atas. Fokus dari revisi ini adalah membuat distraktor yang lebih efektif, sehingga bisa menarik perhatian siswa dengan kemampuan rendah, sementara tetap dikenali sebagai jawaban yang salah oleh siswa dengan kemampuan tinggi.
3. Soal yang Gagal (20,00%): Terdapat 3 soal (nomor 5, 9, dan 12) yang masuk dalam kategori Tidak Memuaskan (D < 0,20). Soal-soal ini gagal membedakan antara kelompok siswa. Artinya, siswa yang cerdas dan siswa yang kurang cerdas memiliki kesempatan yang hampir sama untuk menjawab butir soal ini, atau bahkan lebih banyak siswa dari kelompok bawah yang memberikan jawaban yang benar. Pada-pada soal tersebut, siswa dari dua kelompok memiliki kesempatan yang hampir serupa untuk menjawab dengan tepat, bahkan dalam beberapa kasus siswa dari kelompok yang lebih rendah memberikan jawaban benar lebih banyak. Oleh karena itu, butir soal ini wajib direvisi total atau dihilangkan karena tidak memenuhi standar kualitas tes.

Dari hasil yang disebutkan, dapat disimpulkan bahwa mutu soal ujian akhir Fisika, jika dilihat dari aspek daya pembeda, menunjukkan kualitas yang cukup memuaskan, sesuai dengan penelitian sebelumnya (Irawati et al, 2020) tentang hasil ulangan Fisika mendapat 65% soal efektif (50% sangat baik +15% baik), 5% marginal (cukup baik), dan 30% gagal (tidak baik, nomor 1,3,7). Hasil hampir mirip dengan penelitian ini, di mana 66,67% soal efektif (10 butir baik sekali + baik), 13,33% marginal (nomor 1,2,8), dan 20% gagal (nomor 5,9,12). Distribusi ini memperkuat bahwa proporsi soal efektif 60-70% sudah mencukupi standar tes fisika SMA. Kesamaan dalam pola ini mengindikasikan bahwa secara empiris, instrumen tes Fisika di tingkat SMA cenderung memiliki angka soal efektif yang lebih tinggi, meskipun masih ada beberapa butir yang perlu diperbaiki atau dihapus.

Temuan ini juga sejalan dengan studi-studi lain dalam evaluasi pembelajaran Fisika yang menunjukkan bahwa persentase soal dengan daya pembeda yang baik pada rentang 60 hingga 70 persen sudah dianggap cukup untuk penilaian hasil belajar di tingkat

SMA. Penelitian oleh (Pradita et al., 2023) dikatakan bahwa penting untuk mengenali siswa yang telah menguasai pelajaran dibandingkan dengan yang belum, ketika menggunakan instrumen yang memiliki kemampuan membedakan yang semakin baik. Daya pembeda yang efektif menunjukkan bahwa soal dapat menyoroti perbedaan tingkat kemampuan berpikir siswa, terutama dalam hal pemahaman konsep, penerapan prinsip, serta logika ilmiah (Nurhalimah et al., 2022).

Dengan demikian, distribusi daya pembeda dalam penelitian ini menegaskan bahwa alat ukur yang digunakan sudah memenuhi syarat minimum kualitas untuk ujian Fisika di SMA. Meskipun demikian, adanya soal yang marginal dan gagal masih memerlukan perhatian untuk dilakukan revisi terhadap pilihan jawaban yang salah atau perbaikan dalam pembuatan soal agar kualitas instrumen bisa ditingkatkan dalam evaluasi yang akan datang.

D. KESIMPULAN

Berdasarkan analisis yang telah dilakukan sebelumnya terhadap kualitas soal untuk ujian akhir Fisika kelas X di tahun ajaran 2025/2026 di MAN Purworejo, dapat disimpulkan bahwa instrumen tes menunjukkan validitas yang memadai dan tingkat reliabilitas yang tinggi. Namun, soal-soal tersebut cenderung mudah sehingga daya pembeda perlu ditingkatkan. Soal-soal yang tidak sesuai dengan kriteria validitas dan memiliki daya pembeda yang rendah harus diperbaiki agar ujian dapat menilai kemampuan siswa dengan lebih tepat, dengan perbaikan dilakukan berdasarkan alasan kelemahannya. Setelah ujian dilaksanakan, guru perlu melakukan evaluasi pasca-tes untuk menyempurnakan kualitas instrumen yang telah disusun.

Saran bagi peneliti selanjutnya, disarankan untuk membandingkan seberapa efektif AnBuso dengan perangkat lunak analisis lainnya dan juga untuk meneliti efek pelatihan terhadap peningkatan kualitas alat ukur yang digunakan oleh guru.

DAFTAR PUSTAKA

Adawiah, R., Ruchliyadi, D. A., & Hiliadi, W. (2025). *Pelatihan Analisis Instrumen Penilaian Pembelajaran dengan Aplikasi AnBuso Di SMP Negeri 21 Banjarmasin*. *4*(4), 915–922. https://doi.org/10.60004/komunita.v4i4.330

Farhan et al. (2021). *Peningkatan Kualitas Pembelajaran Melalui Pelatihan Analisis Instrumental Tes Bantuan Software AnBuso*. *04*. https://doi.org/10.29303/jppm.v4i4.2861

Farida, A. M. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, *1*(1).

Fathoni, A. (2023). Menyingkap Peran Penting Evaluasi dalam Pembelajaran. *Jurnal El-Makrifat Jurnal Pendidikan Dan Kajian Keislaman|*, *1*(1), 31–40.

Hartono, I. D. I., Tenriawaru, A. B., & Ningsih, K. (2024). Analisis Butir Soal Penilaian Sumatif Ipa Kelas Vii Smp Negeri 3 Pontianak Menggunakan Anates. *Jurnal Kajian Pembelajaran Dan Keilmuan*, *8*(2), 162–171. https://doi.org/10.26418/jurnalkpk.v8i2.78282

Heriyati et al. (2023). Pemanfaatan Aplikasi AnBuso Dalam Penilaian Hasil Beajar Di SD Negeri Sukamaju Baru 1 Depok. *Jurnal PADI (Pengabdian Masyarakat Dosen Indonesia)*, *6*(2). https://doi.org/DOI: https://doi.org/10.51836/jpadi.v6i2.599

Irawati et al. (2020). Analisis Butir Soal Ujian Akhir Semester Gasal Menggunakan Program Anbuso di SMA Negeri 1 Boyolali. *Jurnal Materi Dan Pembelajaran Fisika*, *10*(1). https://doi.org/DOI : https://doi.org/10.20961/jmpf.v10i1.42084

Maulidah, H., Sukarno, S., & Syefrinando, B. (2023). Analisis Kualitas Instrumen Tes Fisika Kelas X Menggunakan Software Anates. *Physics and Science Education Journal (PSEJ)*, 153–162. https://doi.org/10.30631/psej.v2i3.1660

Nurhalimah, S., Hidayati, Y., Rosidi, I., Madura, U. T., & Soal, K. (2022). Hubungan Antara Validitas Item Dengan Daya Pembeda Dan. *Jurnal Natural Science Educational Research*, *4*(3), 249–257.

Pinaka, & A. (2024). Analisis kualitas

instrumen tes: indikator daya pembeda dalam mengidentifikasi kompetensi siswa. *Jurnal Ilmu Pendidikan Dan Pembelajaran*, *1–10*.

Pradita, E., Megawanti, P., & Yulianingsih. (2023). Analisis Tingkat Kesukaran, Daya Pembeda, dan Fungsi Distraktor PTS Matematika SMPN Jakarta. *Himpunan: Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, *3*(1), 109–118.

Purba, Y. O., Fadhilaturrahmi, F., Purba, J. T., & Siahaan, K. W. A. (2021). Teknik Uji Instrumen Penelitian Pendidikan. In A. Masruroh (Ed.), *CV WIdina Media Utama*. CV WIDINA MEDIA UTAMA.

Rajagukguk, D. N. (2023). (2023). Mampu Memilih Soal Berdasarkan Tingkat Kesukaran. *JurnalPendidikanSosialdanHumaniora*, *2*(4).

Saputri et al. (2023). Analisis Instrume Assesmen: Valditas, Reliabilitas, Tingkat Kesukaran dan Daya Beda Butir Soal. *Jurnal Ilmiah PGSD STKIP Subang*, *09*, 2986–2995.

Surya et al. (2023). Analisis Kualitas Intruemen Tes Hasil Belajar IPAS Materi Wujud Zat dan Prerubahannya. *Jurnal Pendidikan Dasar*, *7*(2).

Yolanda, R. (2023). Analisis butir soal ulangan harian Fisika dengan Anates. *SIPTEK Seminar Nasional*, *1*(1).

Zayrin et al. (2025). Analisis Instrumen Penelitian Pendidikan (Uji Validitas Dan Relibilitas Instrumen Penelitian). *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*, *3*(2), 780–789. https://doi.org/10.61104/jq.v3i2.1070