

---

## ANALISIS BUTIR SOAL PERPAJAKAN PPH PASAL 21 MENGGUNAKAN SOFTWARE ANATES PADA PENDEKATAN TEORI TES KLASIK

Salsabella Nur Ikhsaniyah<sup>1</sup>, Amanda Dwi Kurnia<sup>2</sup>, Maharani Zuroida<sup>3</sup>, Vivi Pratiwi<sup>4</sup>,  
Luqman Hakim<sup>5</sup>

<sup>1</sup> Universitas Negeri Surabaya, salsabella.23165@mhs.unesa.ac.id

<sup>2</sup> Universitas Negeri Surabaya, amanda.23113@mhs.unesa.ac.id

<sup>3</sup> Universitas Negeri Surabaya, maharani.23089@mhs.unesa.ac.id

<sup>4</sup> Universitas Negeri Surabaya, vivipratiwi@unesa.ac.id

<sup>5</sup> Universitas Negeri Surabaya, luqmanhakim@unesa.ac.id

---

### Abstrak

Kualitas butir soal menjadi instrumen penting untuk memastikan keberhasilan evaluasi pembelajaran. Evaluasi yang baik tidak hanya mencerminkan pencapaian siswa, tetapi juga menjadi alat refleksi bagi pendidik untuk memperbaiki proses pembelajaran. Dalam konteks pembelajaran Pajak Penghasilan (PPH) Pasal 21, tantangan penyusunan soal semakin kompleks karena sifat materi perpajakan yang terus mengalami perubahan regulasi. Penelitian ini bertujuan untuk menganalisis kualitas 15 butir soal PPh Pasal 21 menggunakan TER (Tarif Efektif Rata-Rata) pada siswa kelas XI SMKS Ketintang Surabaya menggunakan pendekatan Teori Tes Klasik melalui software Anates. Analisis dilakukan pada aspek validitas, reliabilitas, daya beda, dan tingkat kesukaran soal. Metode yang digunakan dalam penelitian ini adalah metode survey dengan pendekatan deskriptif kuantitatif. Hasil penelitian menunjukkan bahwa 9 butir soal dinyatakan layak digunakan karena memenuhi kriteria evaluasi yang valid dan reliabel, sementara 6 butir soal memerlukan perbaikan lebih lanjut. Soal-soal yang memerlukan revisi memiliki kelemahan pada validitas rendah, daya beda yang buruk, atau tingkat kesukaran yang tidak sesuai. Penelitian ini memberikan kontribusi dalam meningkatkan kualitas soal berbasis HOTS agar dapat mengukur kemampuan siswa secara lebih objektif, akurat, dan relevan dengan kebutuhan pembelajaran.

**Kata Kunci:** *Butir Soal, Evaluasi, Perpajakan, Soal HOTS, Software Anates*

---

### Abstract

Item quality is an important instrument to ensure the success of learning evaluation. A good evaluation not only reflects student achievement, but also becomes a reflection tool for educators to improve the learning process. In the context of learning Income Tax (ITA) Article 21, the challenge of preparing questions is increasingly complex due to the nature of taxation material that continues to undergo regulatory changes. This study aims to analyze the quality of 15 items of Income Tax Article 21 using TER (Average Effective Rate) on grade XI students of SMKS Ketintang Surabaya using the Classical Test Theory approach through Anates software. The analysis was conducted on the aspects of validity, reliability, differentiation, and difficulty level of the questions. The results showed that 9 items were declared suitable for use because they met the criteria for valid and reliable evaluation, while 6 items required

further improvement. The items that required revision had weaknesses in low validity, poor differentiation, or inappropriate difficulty levels. This research contributes to improving the quality of HOTS-based questions in order to measure students' abilities more objectively, accurately, and relevant to learning needs. Suggestions for future research can integrate other approaches besides using Classical Test Theory, such as Item Response Theory (IRT), to obtain a more in-depth analysis of item quality.

**Keywords:** Question Items, Evaluation, Taxation, HOTS Questions, Anates Software

\*✉ Corresponding author:

**Amanda Dwi Kurnia**

amanda.23113@mhs.unesa.ac.id

Article History:

Received Des 1, 2024

Revised Des 9, 2024

Accepted Des 12, 2024

## PENDAHULUAN

Dalam dunia pendidikan, kualitas pembelajaran tidak hanya diukur dari bagaimana materi disampaikan, tetapi juga dari bagaimana hasil belajar siswa dievaluasi. Evaluasi pembelajaran merupakan komponen penting dalam mengukur capaian kompetensi peserta didik dan efektivitas proses pembelajaran (Bici & Çela, 2017; Suardipa, 2023; Wisman et al., 2021). Evaluasi yang baik dapat memberikan gambaran akurat terkait keberhasilan peserta didik dalam memahami dan menerapkan materi yang telah diajarkan (Ilham Hidayat et al., 2024a; Sohilit & Asep, 2023). Salah satu instrumen evaluasi yang digunakan adalah butir soal, yang memiliki peran penting dalam memastikan ketercapaian hasil belajar, sehingga diperlukan analisis mendalam untuk memastikan butir soal mampu mengukur kompetensi peserta didik secara akurat. Namun seiring perkembangan zaman, siswa masa kini tidak hanya dituntut untuk menghafal, tetapi juga harus berpikir kritis, menganalisis permasalahan yang kompleks, dan menemukan solusi yang kreatif (Fikri et al., 2023; Ilham Hidayat et al., 2024b).

Namun, dalam implementasinya soal berbasis HOTS dalam pembelajaran

masih sering kali belum mampu memenuhi kriteria evaluasi yang valid dan reliabel (Fikri et al., 2023), dimana ditemukan butir soal tidak memenuhi standar kualitas, baik dari segi validitas, reliabilitas, tingkat kesukaran, daya pembeda, maupun efektivitas pengecoh. Masalah ini dapat mengakibatkan hasil evaluasi yang dihasilkan kurang mencerminkan kemampuan peserta didik. Dalam pembelajaran materi Pajak Penghasilan (PPh) Pasal 21, tantangan dalam menyusun soal menjadi lebih kompleks karena sifat materi perpajakan yang tergolong sulit. Hal ini disebabkan oleh adanya beragam aturan yang bersifat teknis dan dinamis, serta seringnya terjadi perubahan regulasi perpajakan. Kondisi ini membuat siswa tidak hanya dituntut memahami konsep, tetapi juga mampu mengaplikasikan materi sesuai dengan aturan yang berlaku, dan mampu menganalisis skenario perpajakan seperti perhitungan pajak bagi pegawai dengan jenis pekerjaan berbeda.

Penelitian ini bertujuan untuk menganalisis kualitas butir soal PPh Pasal 21 dengan menggunakan pendekatan Teori Tes Klasik melalui software Anates, yang diaplikasikan pada siswa kelas XI di SMKS Ketintang Surabaya dengan melibatkan sebanyak 45 siswa sebagai subjek penelitian. Pendekatan teori tes klasik

menjadi salah satu metode yang banyak digunakan untuk menilai kualitas butir soal, terutama pada aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh (A Irwan; APS Kerans, 2022; Lestari M, 2023) . Pendekatan ini memberikan hasil analisis yang rinci, berfungsi sebagai dasar untuk perbaikan dalam instrumen, dan mengidentifikasi kelemahan pada butir soal sehingga nantinya dapat diperbaiki dan soal yang dihasilkan mampu mengukur kemampuan siswa secara lebih akurat dan obyektif (Aryani et al., 2021)

Berbagai penelitian terdahulu juga menekankan pentingnya analisis butir soal dalam meningkatkan mutu evaluasi pembelajaran seperti penelitian yang dilakukan oleh Azhriel et al., (2024) Sari & Yudha, (2020) menunjukkan bahwa software seperti Anates dapat membantu dalam menganalisis kelemahan soal secara cepat dan tepat. Selain itu, penelitian yang dilakukan oleh Aryani et al., (2021) menunjukkan bahwa pendekatan kuantitatif melalui teori tes klasik memberikan informasi rinci tentang mutu setiap soal, yang berkontribusi terhadap perbaikan instrumen pembelajaran. Maka dari itu rumusan masalah dalam penelitian ini adalah:

1. Bagaimana kualitas butir soal PPh Pasal 21 pada siswa kelas XI SMKS Ketintang Surabaya berdasarkan pendekatan teori tes klasik menggunakan Anates?
2. Apa saja rekomendasi perbaikan yang dapat diimplementasikan berdasarkan hasil analisis?

Manfaat dari penelitian ini diharapkan dapat membantu guru dalam meningkatkan kualitas materi evaluasi pembelajaran khususnya pada materi PPh Pasal 21. Hasil penelitian ini juga dapat menjadi gambaran dalam penyusunan soal yang lebih berkualitas, lebih relevan, dan sesuai dengan kebutuhan siswa. Selain itu,

diharapkan penelitian ini dapat memberikan kontribusi metodologis melalui penggunaan perangkat lunak Anates dalam analisis soal, yang dapat diterapkan pada bidang lain. Keterbaruan penelitian ini terletak pada integrasi antara analisis kualitas soal berbasis HOTS dan konteks materi perpajakan, khususnya PPh Pasal 21. Selain itu, penelitian ini mengangkat pentingnya analisis mendalam terhadap kualitas butir soal, yang sering kali diabaikan dalam praktik pembelajaran. Penelitian ini diharapkan dapat menjadi langkah strategis untuk mendukung penyusunan evaluasi pembelajaran yang lebih berkualitas dan relevan dengan perkembangan regulasi perpajakan di Indonesia.

## METODE PENELITIAN

Penelitian ini menggunakan metode survei dengan pendekatan deskriptif kuantitatif yang bertujuan untuk menggambarkan karakteristik soal pilihan ganda dengan tingkat kesulitan HOTS (Higher Order Thinking Skills) pada materi Pajak Penghasilan Pasal 21 TER. Soal yang digunakan sebanyak 15 butir pilihan ganda, masing-masing dengan lima pilihan jawaban (A, B, C, D, E), yang disebarakan kepada siswa melalui Google Form. Subjek penelitian terdiri dari 45 siswa kelas XI jurusan Akuntansi dan Keuangan Lembaga (AKL) di SMKS Ketintang Surabaya, dengan rincian 15 siswa (semua perempuan) dari kelas XI AKL 1 dan 30 siswa (24 perempuan dan 6 laki-laki) dari kelas XI AKL 2.



## 1. Skor Data Dibobot

No Urut	No Subyek	Kode>Nama	Benar	Salah	Kosong	Skor Asli	Skor Bobot
1	1	Niken Kaila Rifaza	8	7	0	8	24
2	2	Melinda Putri Cahyani	1	14	0	1	3
3	3	Miftahul Ana Isa Ardha	3	12	0	3	9
4	4	Rosyifa Ferdiana	3	12	0	3	9
5	5	Muhammad Ruben	9	6	0	9	27
6	6	Irfangga Atha Murad Eshan	13	2	0	13	39
7	7	Damar wahdatul Robbi	8	7	0	8	24
8	8	Andini Rayya Safabila	6	9	0	6	18
9	9	Annisia Wea Imanito	5	10	0	5	15
10	10	Ayu Mazidatul Ilmiah	7	9	0	7	21
11	11	Karina Gladish Atandi	7	9	0	7	21
12	12	Sofia Ulfa A	8	7	0	8	24
13	13	Agry vanella t	7	9	0	7	21
14	14	Widya Nur Oktavia	3	12	0	3	9
15	15	Safiatunnisa	3	12	0	3	9
16	16	Bella Dwi aminanti	3	12	0	3	9
17	17	Intan Karisma	7	9	0	7	21
18	18	Naura ardisalya	9	6	0	9	27
19	19	Ikhlama Ratu Alifa	9	6	0	9	27
20	20	Anggun Cahyani	9	6	0	9	27
21	21	Chika Shafa Aliza	9	6	0	9	27
22	22	Abidah Ardelia Feldawati	9	6	0	9	27
23	23	Verita Amelia Sari	3	12	0	3	9
24	24	Regina Ananta Sari	9	6	0	9	27
25	25	hammad Wisnu Agung Prat	9	6	0	9	27
26	26	Azalia Biyan Octivia	9	6	0	9	27
27	27	Radian Abiyasa	8	7	0	8	24
28	28	Putri Ayunda Salsabila	6	9	0	6	18

Sumber : Diolah oleh Penulis (2024)  
Gambar 2. Skor Data Dibobot

Data hasil yang didapatkan dari analisis soal dengan software Anates v4 pada Gambar 2 di skor data dibobot yaitu mendapatkan hasil dengan jumlah subyek 45 peserta didik Kelas XI AK-1 dan XI AK-2 SMKS Ketintang, serta jumlah butir soal terdapat 15 soal, dan dengan 5 buah pilihan jawaban terdiri dari A, B, C, D, dan E. Skor yang akan didapatkan peserta didik saat menjawab dengan benar yaitu 3 poin dan jika peserta didik menjawab dengan salah maka akan mendapatkan skor 0 di setiap soal pilihan ganda yang dikerjakan.

## 2. Koefisien Reliabilitas

Tabel 1. Koefisien Reliabilitas

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

Uji reliabilitas memiliki tujuan untuk mengetahui seberapa instrumen dapat

Reliabilitas Tes

Rata2=6.22 Simpang Baku= 2.49 KorelasiXY= 0.52 Reliabilitas Tes = 0.68

No. Urut	No. Subyek	Kode>Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	1	Niken Kaila Rifaza	5	3	8
2	2	Melinda Putri Cahyani	0	1	1
3	3	Miftahul Ana Isa Ardha	2	1	3
4	4	Rosyifa Ferdiana	1	2	3
5	5	Muhammad Ruben	4	5	9
6	6	Irfangga Atha Murad Eshan	7	6	13
7	7	Damar wahdatul Robbi	4	4	8
8	8	Andini Rayya Safabila	2	4	6
9	9	Annisia Wea Imanito	2	3	5
10	10	Ayu Mazidatul Ilmiah	3	4	7
11	11	Karina Gladish Atandi	4	3	7
12	12	Sofia Ulfa A	4	4	8
13	13	Agry vanella t	4	3	7
14	14	Widya Nur Oktavia	3	0	3
15	15	Safiatunnisa	3	0	3
16	16	Bella Dwi aminanti	3	0	3
17	17	Intan Karisma	4	3	7
18	18	Naura ardisalya	4	5	9
19	19	Ikhlama Ratu Alifa	4	5	9
20	20	Anggun Cahyani	4	5	9
21	21	Chika Shafa Aliza	4	5	9
22	22	Abidah Ardelia Feldawati	4	5	9
23	23	Verita Amelia Sari	2	1	3
24	24	Regina Ananta Sari	4	5	9
25	25	hammad Wisnu Agung Prat	4	4	8
26	26	Azalia Biyan Octivia	4	4	8
27	27	Radian Abiyasa	4	4	8
28	28	Putri Ayunda Salsabila	2	4	6

konsisten dan dipercaya. (Fietri et al., 2021). Menurut Syahril et al. (2019) Reliabilitas mengacu pada keyakinan bahwa suatu instrumen memiliki tingkat kepercayaan yang memadai untuk digunakan sebagai alat dalam mengumpulkan data, karena instrumen tersebut dianggap cukup andal.

Sumber : Diolah oleh Penulis (2024)

Gambar 3. Koefisien Reliabilitas

Pada gambar 3 didapatkan skor data pada gambar 3 yang menampakkkan hasil rata-rata peserta didik yaitu mendapatkan 6.22

dan standar deviasi 2.49. Uji reliabilitas perlu digunakan dan dipertimbangkan karena pada tes ini hasilnya cukup andal dan terpercaya untuk hasil sebuah tes. Hasil tes reliabilitas menggunakan software Anates v4 menunjukkan besar nilai rata – rata, simpangan baku, korelasi, dan nilai reliabilitas dengan 15 butir soal dan didapatkan korelasi XY adalah 0.52 serta reliabilitasnya berada di angka 0.68 yang artinya untuk 15 soal ini cukup baik. Indeks pada koefisien reliabilitas mulai dari -0,061 yang berarti reliabilitas sangat rendah dan yang paling tinggi di indeks 0,737.

### 3. Kelompok Unggul dan Asor

Kelompok Unggul																		
No. Umat	No Subyek	Kode/Nama Subyek	Skor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	6	Eterangga Alha Masud Eshah	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	5	Muhammad Ruben	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	18	Naura ardasya	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	19	Ikhina Ratu Alifa	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	20	Anggun Cahyani	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	21	Chita Shafa Alha	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	22	Abdich Ardika Faldawati	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	24	Regina Aranta Sari	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	Niken Kaba Rizka	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

  

Kelompok Asor																		
No. Umat	No Subyek	Kode/Nama Subyek	Skor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	36	Chalitta Nuryana Az-zahra	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	37	Aina Marsha Aurelia	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	39	Jendra Fatma Salsita Dewanti	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	3	Mitahul Anis Ica Andra	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	4	Rozalia Ferdiana	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	14	Widya Nur Octavia	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	15	Safarunnisa	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	16	Bella Dwi ammani	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	23	Venika Amelia Sari	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	38	Aprilia Wulandari	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	42	Aina Imawati	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	2	Melinda Putri Cahyani	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Jml Zeb Siswa			4	2	2	4	3	2	6	1	1	1	4	0	0	0	0	0

Sumber : Diolah oleh Penulis (2024)

Gambar 4. Kelompok Unggul dan Asor

Pada software Anates v4 subjek dikelompokkan berdasarkan kelompok unggul dan kelompok asor. Kelompok unggul terdiri dari siswa yang memiliki skor yang tinggi, sedangkan kelompok asor adalah bagian siswa yang mendapatkan skor rendah (Cahyaningrum et al., 2023).

Pada Gambar 4 terdapat 12 peserta didik masuk ke kelompok unggul dengan Erlangga yang mendapatkan skor paling tinggi yaitu 13 dan pada kelompok asor juga terdapat 12 peserta didik yang mendapatkan skor rendah dengan Melinda yang hanya mendapatkan skor 1.

### 4. Daya Pembeda Butir Soal

Tabel 2. Indeks Daya Pembeda

Indeks	Kriteria
0,70 – 1,00	Sangat Baik (Soal Diterima)
0,40 – 0,70	Baik (Soal Diterima)
0,20 – 0,40	Cukup (Soal Diterima dan Diperbaiki)
0,00 – 0,20	Jelek (Soal Diperbaiki)
Negatif	Soal Dibuang

Daya pembeda merujuk pada kemampuan sebuah butir soal untuk membedakan hasil tes berdasarkan jawaban yang benar dan salah. Secara sederhana, daya pembeda juga diartikan sebagai kemampuan soal untuk mengidentifikasi perbedaan antara siswa dengan kemampuan tinggi dan siswa dengan kemampuan rendah (Akhmadi, 2021).

Daya Pembeda					
Kembali Ke Menu Utama Cetak					
Jml Subyek= 45		Klp atas/bawah (n) = 12		Butir Soal = 15	
No Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	1	12	4	8	66,67
2	2	11	2	9	75,00
3	3	4	2	2	16,67
4	4	5	4	1	8,33
5	5	3	3	0	0,00
6	6	11	2	9	75,00
7	7	11	6	5	41,67
8	8	2	1	1	8,33
9	9	12	1	11	91,67
10	10	11	1	10	83,33
11	11	9	4	5	41,67
12	12	8	0	8	66,67
13	13	0	0	0	0,00
14	14	8	5	3	25,00
15	15	1	0	1	8,33

## Sumber : Diolah oleh Penulis (2024)

Gambar 5. Tes Daya Pembeda Butir Soal

Selanjutnya, tes daya pembeda butir soal bisa dilihat di Gambar 5 digunakan untuk mengetahui hasil tes dengan hasil jawaban yang salah dan dari 15 butir soal dan subyek 45 peserta didik didapatkan indeks 0,00 (Soal jelek dan perlu diperbaiki) - 0,91 (Soal sangat baik dan soal diterima serta digunakan). Dari 15 soal terdapat 2 soal mudah, 9 soal sedang, 1 soal sukar dan 3 soal sangat sukar. Pada menu software Anates v4 juga terdapat menu korelasi butir dengan skor total yang dapat digunakan untuk mengetahui seberapa signifikan soal.

## 5. Tingkat Kesukaran Butir Soal

Tabel 3. Indeks Tingkat Kesukaran

Indeks	Tingkat Kesukaran
0,00 – 0,30	Sangat Baik
0,30 – 0,70	Sedang
0,70 – 1,00	Mudah

Soal yang baik adalah soal yang memiliki tingkat kesulitan yang seimbang, tidak terlalu mudah maupun terlalu sulit. Soal yang terlalu mudah cenderung tidak mampu mendorong peserta didik untuk mengasah kemampuan mereka dalam menyelesaikan soal. Sebaliknya, soal yang terlalu sulit berisiko membuat peserta didik merasa putus asa karena dianggap berada di luar kemampuan mereka (Kaka et al., 2024). Oleh karena itu, tingkat kesukaran (TK) soal merupakan ukuran yang menunjukkan sejauh mana soal tersebut sulit untuk dikerjakan. Sebuah soal dikatakan berkualitas jika memiliki tingkat kesulitan yang proporsional, yakni tidak terlalu mudah maupun terlalu sulit.

Tingkat Kesukaran

Tingkat Kesukaran [Kembali Ke Menu Utama](#) [Cetak](#)

Jml Subyek = 45 Butir Soal = 15

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)	Tafsiran
1	1	33	73,33	Mudah
2	2	34	75,56	Mudah
3	3	24	53,33	Sedang
4	4	23	51,11	Sedang
5	5	16	35,56	Sedang
6	6	28	62,22	Sedang
7	7	23	51,11	Sedang
8	8	6	13,33	Sangat Sukar
9	9	23	51,11	Sedang
10	10	25	55,56	Sedang
11	11	15	33,33	Sedang
12	12	10	22,22	Sukar
13	13	5	11,11	Sangat Sukar
14	14	14	31,11	Sedang
15	15	1	2,22	Sangat Sukar

## Sumber : Diolah oleh Penulis (2024)

Gambar 6. Tingkat Kesukaran Soal

Hasil yang dapat diperoleh dari menggunakan software Anates v4 dari ke lima belas soal terdapat soal yang mudah hingga sulit. Dengan menganalisis indeks tingkat kesukaran dapat diketahui soal tersebut mudah, sedang atau sangat sukar, seperti pada soal nomor satu mendapatkan skor 0,73 yang berarti memiliki tingkat kesukaran yang mudah.

## 6. Korelasi Butir Soal

Tabel 4. Indeks Kriteria Validitas

Kriteria	Validitas
0,800 – 1,00	Validitas Soal Sangat Tinggi
0,600 – 0,800	Validitas Soal Tinggi
0,400 – 0,600	Validitas Soal Cukup
0,200 – 0,400	Validitas Soal Rendah
0,00 – 0,200	Validitas Soal Sangat Rendah

Untuk memperoleh data yang valid, instrumen yang digunakan untuk penilaian juga harus valid. Validitas suatu alat ukur menunjukkan sejauh mana alat ukur itu mengukur apa yang seharusnya diukur.

Hasil yang didapatkan dilihat dari Gambar 7, yaitu 3 soal signifikan dan 3 soal sangat signifikan, sisanya validitas soal yang masih rendah.

Korelasi Butir dengan Skor Total

Korelasi Skor Butir dg Skor Total [Kembali Ke Menu Utama](#) [Cetak](#)

Jml Subyek= 45 Butir Soal = 15 [Info tentang batas signifikansi](#)

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0,543	Signifikan
2	2	0,659	Sangat Signifikan
3	3	0,211	-
4	4	0,160	-
5	5	0,178	-
6	6	0,553	Signifikan
7	7	0,286	-
8	8	0,097	-
9	9	0,737	Sangat Signifikan
10	10	0,661	Sangat Signifikan
11	11	0,338	-
12	12	0,559	Signifikan
13	13	-0,061	-
14	14	0,212	-
15	15	0,414	-

Sumber : Diolah oleh Penulis (2024)

Gambar 7. Tes Korelasi Butir dengan Skor Total

### 7. Kualitas Pengecoh

Kemungkinan-kemungkinan jawaban yang dipasangkan pada setiap butir jawaban pada soal pilihan ganda yang digunakan untuk menggambarkan pola disebut distraktor (Akhmadi, N. M., 2021). Lalu, kualitas pengecoh soal disini digunakan untuk melihat seberapa bagus untuk pilihan jawaban yang diberikan untuk soal tersebut.

Tabel 5. Indeks Kualitas Pengecoh

Tanda	Keterangan
**	Kunci Jawaban
++	Sangat Baik
+	Baik
-	Kurang Baik
--	Buruk
---	Sangat Buruk

Pada Gambar 8 tersebut dapat dilihat ada kotak yang diberikan warna merah, yang dapat diartikan pilihan jawaban tersebut itu masih kurang baik dan perlu diperbaiki seperti untuk soal nomor 1 itu untuk jawaban c dan d itu perlu direvisi.

Kualitas Pengecoh

Jml Subyek= 45 Butir Soal = 15 \*\* : Kunci Jawaban + : Baik -- : Buruk  
++ : Sangat Baik - : Kurang --- : Sangat Buruk

Kualitas Pengecoh [Kembali Ke Menu Utama](#) [Cetak](#)

No Butir Baru	No Butir Asli	a	b	c	d	e	*
1	1	33**	4+	0-	7--	1-	0
2	2	5--	34**	6--	0-	0-	0
3	3	2-	1--	14--	4++	24**	0
4	4	1--	23**	3+	3+	15--	0
5	5	3-	9++	6++	16**	11-	0
6	6	28**	5++	4++	5++	3+	0
7	7	2-	14--	6++	23**	0-	0
8	8	6+	2-	6**	21--	10++	0
9	9	8+	7+	23**	6++	1-	0
10	10	25**	10--	3+	1-	6++	0
11	11	19--	15**	4+	5+	2-	0
12	12	5+	25--	4-	10**	1-	0
13	13	10++	20--	4-	6+	5**	0
14	14	12-	14**	18--	1-	0-	0
15	15	9++	19-	15+	1-	1**	0

Sumber : Diolah oleh Penulis, 2024

Gambar 8. Tes Kualitas Pengecoh 8. Rekap Analisis Butir

FRekapButir

Rekap Analisis Butir [Kembali Ke Menu Utama](#) [Cetak](#)

Rate=6.22 Simpang Baku= 2.49 Korelasi<sup>2</sup>Y<sup>2</sup>= 0.52 Reliabilitas Tes = 0.68 Butir Soal = 15

Btr Baru	Btr Asli	D. Pembeda(%)	T. Kesukaran	Korelasi	Sign. Korelasi
1	1	66.67	Mudah	0,543	Signifikan
2	2	75.00	Mudah	0,659	Sangat Signifikan
3	3	16.67	Sedang	0,211	-
4	4	8.33	Sedang	0,160	-
5	5	0.00	Sedang	0,178	-
6	6	75.00	Sedang	0,553	Signifikan
7	7	41.67	Sedang	0,286	-
8	8	8.33	Sangat Sukar	0,097	-
9	9	91.67	Sedang	0,737	Sangat Signifikan
10	10	83.33	Sedang	0,661	Sangat Signifikan
11	11	41.67	Sedang	0,338	-
12	12	66.67	Sukar	0,559	Signifikan
13	13	0.00	Sangat Sukar	-0,061	-
14	14	25.00	Sedang	0,212	-
15	15	8.33	Sangat Sukar	0,414	-

Sumber : Diolah oleh Penulis (2024)

Gambar 9. Rekap Analisis Butir

Kegiatan analisis yang telah dilaksanakan di atas merupakan analisis yang dilakukan secara menyeluruh, yaitu analisis skor data dibobot, koefisien reliabilitas, kelompok unggul dan asor, daya pembeda butir soal, tingkat kesukaran butir soal, korelasi butir soal, dan kualitas pengecoh soal. Analisis

yang terakhir adalah analisis rekap butir soal secara keseluruhan, sehingga butir soal yang nantinya akan digunakan atau akan direvisi bisa terlihat. Kegunaan analisis butir soal bukan hanya terbatas untuk meningkatkan butir soal, tetapi juga bermanfaat sebagai dasar peningkatan pembelajaran di kelas secara umum dan untuk meningkatkan keterampilan pada konstruksi tes.

### 9. Rangkuman Hasil Analisis Butir

No Butir Soal	Kriteria Validitas	Koefisien Reliabilitas	Kriteria Daya Beda Soal	Tingkat Kesukaran Soal	Kesimpulan
1	Signifikan	0,543 (Cukup)	0,66 (Sangat Baik)	Mudah	Soal Digunakan
2	Sangat Signifikan	0,659 (Cukup)	0,75 (Baik)	Mudah	Soal Digunakan
3	-	0,211 (Rendah)	0,16 (Jelek)	Sedang	Soal Diperbaiki
4	-	0,160 (Sangat Rendah)	0,8 (Jelek)	Sedang	Soal Diperbaiki
5	-	0,178 (Sangat Rendah)	0,00 (Jelek)	Sedang	Soal Diperbaiki
6	Signifikan	0,553 (Cukup)	0,75 (Sangat Baik)	Sedang	Soal Digunakan
7	-	0,296 (Rendah)	0,41 (Baik)	Sedang	Soal Digunakan
8	-	0,097 (Sangat Rendah)	0,8 (Jelek)	Sangat Sukar	Soal Diperbaiki
9	Sangat Signifikan	0,737 (Tinggi)	0,91 (Sangat Baik)	Sedang	Soal Digunakan
10	Sangat Signifikan	0,661 (Cukup)	0,83 (Sangat Baik)	Sedang	Soal Digunakan
11	-	0,338 (Rendah)	0,41 (Baik)	Sedang	Soal Digunakan
12	Signifikan	0,559 (Cukup)	0,66 (Baik)	Sukar	Soal Digunakan
13	-	-0,061 (Sangat Rendah)	0,00 (Jelek)	Sangat Sukar	Soal Diperbaiki
14	-	0,212 (Rendah)	0,25 (Cukup)	Sedang	Soal Diperbaiki
15	-	0,414 (Rendah)	0,8 (Jelek)	Sangat Sukar	Soal Diperbaiki

Sumber : Diolah oleh Penulis (2024)

Gambar 10. Rangkuman Hasil Analisis Butir

Melalui analisis butir soal, semua komponen pembelajaran dapat diketahui apakah dapat berfungsi sebagaimana mestinya atau tidak, guru juga dapat mengetahui tingkat kemampuan peserta didik dan mendapat gambaran mengenai efektivitas proses pembelajaran serta dapat memutuskan tindak lanjut seperti apa yang akan diterapkan para peserta didik yang mendapat hasil belajar baik maupun kurang baik (Mania et al., 2024). Hasil analisis terhadap lima belas butir soal menunjukkan variasi kualitas soal berdasarkan kriteria validitas, reliabilitas, daya beda, dan tingkat kesukaran. Butir soal 1, 2, 6, 9, 10, dan 12

dinyatakan layak digunakan tanpa perbaikan karena memiliki validitas signifikan sampai dengan sangat signifikan, reliabilitas cukup hingga tinggi (berkisar 0,543 hingga 0,737), daya beda baik hingga sangat baik, serta tingkat kesukaran yang sesuai (mudah hingga sedang).

Sebaliknya, butir soal 3, 4, 5, 11, 13, 14, dan 15 membutuhkan perbaikan karena kelemahan pada beberapa aspek. Sebagian besar butir ini menunjukkan reliabilitas rendah hingga sangat rendah (di bawah 0,338) dan daya beda yang kurang baik hingga sangat buruk (di bawah 0,41), meskipun tingkat kesukarannya bervariasi dari sedang hingga sangat sukar.

Secara khusus, butir soal 7 dan 8 meskipun memiliki validitas signifikan, tetap perlu ditinjau karena daya beda yang cukup hingga baik masih dapat ditingkatkan untuk hasil yang lebih optimal. Analisis ini secara keseluruhan mengindikasikan bahwa sebagian besar soal memerlukan perbaikan, tetapi ada juga sejumlah soal yang sudah memenuhi kriteria kualitas.

### PEMBAHASAN

Berdasarkan hasil penelitian yang telah dilakukan, kualitas butir soal PPH Pasal 21 menunjukkan bahwa dari 15 soal yang dianalisis, sebanyak 9 butir soal dinyatakan layak digunakan karena memenuhi kriteria validitas, reliabilitas, daya beda, dan tingkat kesukaran yang sesuai. Namun, terdapat 6 butir soal yang memerlukan perbaikan karena memiliki kelemahan, seperti validitas yang rendah, daya beda yang kurang baik, atau tingkat kesukaran yang tidak seimbang. Hal ini mengindikasikan bahwa soal yang digunakan sudah cukup konsisten dan dapat dipercaya untuk mengukur kemampuan peserta didik

Hasil penelitian ini didukung oleh kajian sebelumnya, seperti yang dilakukan oleh Sari & Yudha, (2020), yang menunjukkan bahwa penggunaan

perangkat lunak seperti Anates mampu memberikan hasil analisis soal secara cepat dan tepat, sehingga mempermudah identifikasi kelemahan dalam soal. Penelitian oleh Aryani et al., (2021) juga menekankan bahwa pendekatan kuantitatif dengan teori tes klasik mampu memberikan informasi rinci tentang kualitas butir soal, yang sangat berguna untuk memperbaiki mutu instrumen evaluasi pembelajaran. Selain itu, penelitian oleh (Kaka et al., 2024) menyoroti pentingnya mempertahankan keseimbangan tingkat kesukaran soal, karena soal yang terlalu mudah atau terlalu sulit cenderung tidak efektif dalam mengukur kemampuan siswa secara optimal.

Revisi yang diperlukan meliputi peningkatan konstruksi soal, penyesuaian tingkat kesulitan, serta perbaikan pada pengecoh jawaban agar instrumen evaluasi dapat lebih mendukung pengembangan kemampuan siswa secara menyeluruh.

## KESIMPULAN

Penelitian ini menunjukkan bahwa kualitas butir soal berbasis HOTS pada materi PPh Pasal 21 TER yang dikembangkan oleh peneliti dan diujikan kepada siswa kelas XI SMKS Ketintang Surabaya masih bervariasi berdasarkan hasil analisis menggunakan pendekatan Teori Tes Klasik melalui software Anates v4. Dari 15 butir soal yang diujikan, 9 soal dinyatakan layak digunakan karena memenuhi kriteria validitas, reliabilitas, daya beda, dan tingkat kesukaran yang sesuai, sedangkan 6 soal lainnya memerlukan perbaikan. Beberapa kelemahan ditemukan pada soal dengan validitas rendah, daya beda jelek, serta tingkat kesukaran yang tidak proporsional, baik terlalu mudah maupun sangat sukar. Hal ini menunjukkan bahwa butir soal berbasis HOTS membutuhkan perancangan yang lebih matang, terutama untuk materi yang bersifat kompleks seperti PPh Pasal

21. Oleh karena itu, diperlukan revisi pada soal yang belum memenuhi kriteria, seperti memperbaiki konstruksi soal, meningkatkan efektivitas pengecoh, serta menyesuaikan tingkat kesukaran. Penelitian ini diharapkan dapat memberikan wawasan penting dalam menyusun instrumen evaluasi berbasis HOTS yang lebih berkualitas, valid, dan relevan yang sesuai dengan perkembangan regulasi perpajakan sehingga dapat mendukung pembelajaran yang lebih optimal.

## DAFTAR PUSTAKA

- A Irwan; APS Kerans. (2022). Analisis Kualitas Soal Three Tier Test Menggunakan Teori Respon Butir Klasik. <https://doi.org/https://doi.org/10.24252/jpf.v10i2.31068>
- Akhmadi, N. M. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 Sdn Plumbungan Menggunakan Program Anates.
- Aryani, L., Widayat, E., & Sunardjo, S. (2021). Pengaruh Model Pembelajaran Discovery Learning dan Cooperative Learning Tipe Jigsaw terhadap Aktivitas dan Hasil Belajar Siswa. *Edudikara: Jurnal Pendidikan Dan Pembelajaran*, 6(2). <https://doi.org/10.32585/edudikara.v6i2.234>
- Azhriel, M. G., Adam N, R. M., Wiyono, A., & Teknik Otomotif Jl Setiabudhi No, P. (2024). Pembuatan Instrumen Asesmen Pada Pembelajaran Sistem Differential. In Atikanoto: *Journal of Automotive Engineering Education* (Vol. 1, Issue 1).
- Bici, R., & Çela, M. (2017). Education as An Important Dimension of the

- Poverty. *European Journal of Multidisciplinary Studies*, 4(3), 88. <https://doi.org/10.26417/ejms.v4i3.p88-95>
- Akhmadi, M. N. (2021). Analisis Butir Soal Evaluasi Tema 1 Kelas 4 Sdn Plumbungan Menggunakan Program Anates.
- Cahyaningrum, I. Y., Fuady, A., & Sunismi. (2023). Analisis Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran Matematika Kelas Vii Dengan Berbantuan Aplikasi Software Anates. In *Mathema Journal E-Issn* (Vol. 5, Issue 2).
- Fietri, W. A., Zulyusri, & Violita. (2021). Natural Science: Jurnal Penelitian Bidang Ipa Dan Pendidikan Ipa Analisis Butir Soal Biologi Kelas Xi Madrasah Aliyah Sakinah Kerinci Menggunakan Program Komputer Anates 4.0 For Windows (Vol. 7, Issue 1).
- Fikri, M., Munfarida, E., Negeri Saifuddin Zuhri Purwokerto, I. K., Ji Yani No, I. A., & Utara, P. (2023). Konstruksi Berpikir Kritis Dalam Pendidikan Islam: Analisis Tafsir Maudhu'i Berdasarkan Al-Qur'an. *Jurnal Pendidikan Agama Islam Al-Thariqah*, 8(1). [https://doi.org/10.25299/Al-Thariqah.2023.Vol8\(1\).11469](https://doi.org/10.25299/Al-Thariqah.2023.Vol8(1).11469)
- Ilham Hidayat, Muhammad Rizqi Robbani, & Arrasyidin Arrasyidin. (2024a). Strategi Evaluasi Terbaik Menghadapi Kompleksitas Uts Di Lembaga Pendidikan. *Student Scientific Creativity Journal*, 2(1), 165–174. <https://doi.org/10.55606/Sscj-Amik.V2i1.2695>
- Ilham Hidayat, Muhammad Rizqi Robbani, & Arrasyidin Arrasyidin. (2024b). Strategi Evaluasi Terbaik Menghadapi Kompleksitas Uts Di Lembaga Pendidikan. *Student Scientific Creativity Journal*, 2(1), 165–174. <https://doi.org/10.55606/Sscj-amik.v2i1.2695>
- Kaka, L., Oktaviana Bano, V., & Njoeroemana, Y. (2024). Efektivitas Analisis Butir Soal Pilihan Ganda Menggunakan Aplikasi Anates Di Smpn 2 Kanatang. *Jurnal Inovasi Penelitian*, Vol.4 No.9(Vol 4 No 9: Februari 2024), 1–10. <https://doi.org/https://doi.org/10.47492/jip.v4i9.3124>
- Lestari M, (2023) Analisis Tingkat Kesukaran Soal Persamaan Dan Pertidaksamaan Nilai Mutlak Melalui Pendekatan Teori Tes Klasik. (2023). *Indonesian Journal of Learning and Instructional Innovation*, 1(02), 28-33. <https://doi.org/10.20961/ijolii.v1i02.1045>
- Mania, S., Hasrianti, A., Pendidikan Matematika, J., Alauddin Makassar, U., & Corresponding Author, I. (2024). Analisis Kualitas Butir Soal Ulangan Akhir Semester Ganjil Pada Mata Pelajaran Matematika Siswa Kelas X Sma Negeri. *Jurnal Theorems (The Original Reasearch Of Mathematics)*, 8(2).
- Sari, P. I., & Yudha, R. I. (2020). Pemanfaatan Penerapan Media Berbasis Software Anates pada Mata Kuliah Evaluasi Pembelajaran di



- Universitas Batanghari Jambi. *Jurnal Ilmiah Universitas Batanghari Jambi*, 20(1), 81.  
<https://doi.org/10.33087/jiubj.v20i1.732>
- Sohilait, D., & Asep, A. (2023). Evaluasi Program Pembelajaran Berdasarkan Metode EKOP di Sekolah Menengah Atas Negeri 9 Ambon. *Jurnal Ilmiah Profesi Pendidikan*, 8(1b), 539–545.  
<https://doi.org/10.29303/jipp.v8i1b.1247>
- Suardipa, P. (2023). Peran Desain Evaluasi Pembelajaran Untuk Meningkatkan Kualitas Pembelajaran.
- Syahril, Hadiyanto, Adi, N., & Ningrum, Tia Ayu. (2019). Pemberdayaan Aset Sekolah Dan Instansi Pemerintah Melalui Workshop Sistem Manajemen Aset. *1(4)*.  
<https://doi.org/10.5281/zenodo.3522456>
- Lestari M, H. I. T. (2023). *Indonesian Journal of Learning and Instructional Innovation*.  
<https://journal.uns.ac.id/ijolii>
- Wisman, Y., Effrata, E., & Tutesa, T. (2021). Penerapan Konsep Instrumen Evaluasi Hasil Belajar. *Jurnal Ilmiah Kanderang Tingang*, 12(1), 1–9.  
<https://doi.org/10.37304/jikt.v12i1.105>