

ANALISIS KUALITAS SOAL EVALUASI MATA PELAJARAN INFORMATIKA SISWA SMK NEGERI 1 NGAWI MENGGUNAKAN ANATES

Inka Ari Murganingsih
Universitas Negeri Surabaya

Putri Ayu Novita
Universitas Negeri Surabaya

Maulana Aditya Wardana
Universitas Negeri Surabaya

Luqman Hakim
Universitas Negeri Surabaya

Vivi Pratiwi
Universitas Negeri Surabaya

Korespondensi penulis: inka.23214@mhs.unesa.ac.id ; putri.23131@mhs.unesa.ac.id ;
maulana.23166@mhs.unesa.ac.id ; luqmanhakim@unesa.ac.id ; vivipratiwi@unesa.ac.id

Abstract. *This research aims to evaluate the quality of exam questions in the Informatics subject for class XII AKL 4 students at SMKN 1 Ngawi. This evaluation is important to ensure that the questions used can measure student competency validly and reliably. The research method used is quantitative analysis using ANATES software. The data analyzed includes test reliability, discriminating power, level of difficulty, and validity of the questions. The results of the analysis show that the test reliability is classified as very low with a coefficient of 0.20, which indicates that the consistency of question measurement needs to be improved. Based on the difficulty level analysis, the majority of the questions are in the medium category, which shows that the questions are neither too easy nor too difficult for students. However, several questions were found with low or too high discriminating power, so they were less effective in differentiating the abilities of high and low achieving students. In terms of validity, several questions need to be revised to ensure their suitability for learning objectives. The results of this research emphasize the need for improvements in the preparation of exam questions so that they can be used as a better evaluation instrument. Thus, it is hoped that learning evaluation can provide a more accurate picture of student competency achievement. This research also contributes to the development of educational evaluation practices at the vocational high school level.*

Keywords: *Learning Evaluation; Question Analysis; ANATES; Question Quality.*

Abstrak. Penelitian ini bertujuan untuk mengevaluasi kualitas soal ujian mata pelajaran Informatika pada siswa kelas XII AKL 4 di SMKN 1 Ngawi. Evaluasi ini penting untuk memastikan bahwa soal-soal yang digunakan dapat mengukur kompetensi siswa secara valid dan reliabel. Metode penelitian yang digunakan adalah analisis kuantitatif dengan memanfaatkan perangkat lunak ANATES. Data yang dianalisis meliputi reliabilitas tes, daya pembeda, tingkat kesukaran, dan validitas soal. Hasil analisis menunjukkan bahwa reliabilitas tes tergolong sangat rendah dengan koefisien sebesar 0,20, yang mengindikasikan bahwa konsistensi pengukuran soal perlu ditingkatkan. Berdasarkan analisis tingkat kesukaran, mayoritas butir soal berada pada kategori sedang, yang menunjukkan bahwa soal-soal tersebut tidak terlalu mudah maupun terlalu sulit bagi siswa. Namun demikian, ditemukan beberapa soal dengan daya pembeda yang rendah atau

terlalu tinggi, sehingga kurang efektif dalam membedakan kemampuan siswa yang berprestasi tinggi dan rendah. Dari segi validitas, beberapa butir soal perlu direvisi untuk memastikan kesesuaiannya dengan tujuan pembelajaran. Hasil penelitian ini menegaskan perlunya perbaikan dalam penyusunan soal ujian agar dapat digunakan sebagai instrumen evaluasi yang lebih baik. Dengan demikian, diharapkan evaluasi pembelajaran dapat memberikan gambaran yang lebih akurat mengenai pencapaian kompetensi siswa. Penelitian ini juga memberikan kontribusi bagi pengembangan praktik evaluasi pendidikan di tingkat sekolah menengah kejuruan.

Kata Kunci: Evaluasi Pembelajaran; Analisis Soal; ANATES; Kualitas Soal.

LATAR BELAKANG

Evaluasi kualitas soal menjadi salah satu aspek penting dalam pembelajaran karena berperan sebagai alat ukur keberhasilan siswa dalam memahami materi yang diajarkan. Soal ujian yang baik harus mampu memenuhi kriteria validitas, reliabilitas, tingkat kesukaran, dan daya pembeda yang sesuai (Arikunto, 2016). Namun, kenyataan di lapangan menunjukkan bahwa masih banyak soal yang tidak memenuhi kriteria tersebut, sehingga hasil evaluasi pembelajaran kurang mencerminkan kemampuan siswa yang sebenarnya. Hal ini mendorong perlunya penelitian untuk mengidentifikasi dan memperbaiki kualitas instrumen evaluasi, khususnya di tingkat sekolah menengah kejuruan (SMK) yang memiliki peran strategis dalam menyiapkan lulusan dengan kompetensi tertentu.

Penelitian ini berfokus pada evaluasi kualitas soal ujian mata pelajaran Informatika di kelas XII AKL 4 SMKN 1 Ngawi. Informatika sebagai mata pelajaran yang mendukung literasi digital siswa memerlukan instrumen evaluasi yang tidak hanya menilai pengetahuan teoritis tetapi juga kemampuan praktis (Permendikbud No. 37 Tahun 2018). Penelitian ini menggunakan perangkat lunak ANATES untuk menganalisis kualitas soal berdasarkan beberapa parameter, yaitu reliabilitas, daya pembeda, tingkat kesukaran, dan validitas. Penggunaan perangkat lunak ini memungkinkan analisis yang lebih akurat dan sistematis dalam mengevaluasi soal (Nurgiyantoro et al., 2015).

Kajian ini signifikan karena memberikan gambaran tentang kualitas instrumen evaluasi yang digunakan serta memberikan rekomendasi praktis untuk perbaikannya. Selain itu, penelitian ini juga berkontribusi pada pengembangan praktik evaluasi pendidikan di SMK, terutama dalam meningkatkan kompetensi guru dalam menyusun soal yang berkualitas. Artikel ini disusun dengan sistematika sebagai berikut: bagian pertama berisi pendahuluan yang menjelaskan latar belakang, signifikansi, dan tujuan penelitian. Bagian kedua menguraikan metodologi yang digunakan, diikuti dengan hasil

dan pembahasan pada bagian ketiga. Akhirnya, bagian keempat menyajikan kesimpulan dan saran untuk pengembangan lebih lanjut.

Teori dan kajian empiris yang digunakan dalam penelitian ini mencakup prinsip-prinsip evaluasi pembelajaran, teori pengembangan instrumen evaluasi, serta hasil penelitian sebelumnya tentang analisis kualitas soal ujian. Berdasarkan kajian tersebut, penelitian ini berupaya menjawab pertanyaan mengenai kualitas soal ujian yang digunakan dan bagaimana upaya yang dapat dilakukan untuk meningkatkan mutu instrumen tersebut. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan manfaat bagi praktisi pendidikan dalam meningkatkan efektivitas proses evaluasi pembelajaran.

KAJIAN TEORITIS

Kualitas soal ujian merupakan faktor penting dalam menilai kemampuan siswa dalam memahami materi pembelajaran. Soal evaluasi yang baik harus memenuhi beberapa kriteria, antara lain:

- **Validitas:** Soal harus mengukur kemampuan atau pengetahuan yang benar-benar menjadi tujuan pembelajaran.
- **Reliabilitas:** Soal harus memberikan hasil yang konsisten jika diujikan dalam kondisi yang sama.
- **Daya Pembeda:** Soal harus dapat membedakan kemampuan siswa yang memiliki tingkat pemahaman yang berbeda.
- **Tingkat Kesukaran:** Soal harus memiliki tingkat kesulitan yang sesuai dengan tingkat kemampuan siswa.
- **Format Soal:** Jenis soal (pilihan ganda, uraian, isian) juga mempengaruhi kualitas soal dalam menilai kemampuan siswa.

Evaluasi pembelajaran merupakan proses untuk mengumpulkan informasi mengenai pencapaian tujuan pembelajaran dan tingkat penguasaan siswa terhadap materi yang diajarkan. Dalam konteks pendidikan informatika di SMK, evaluasi bertujuan untuk mengetahui sejauh mana siswa menguasai keterampilan dan pengetahuan dalam bidang informatika, yang meliputi teori dan praktik teknologi informasi. Evaluasi ini dapat dilakukan melalui ujian, tugas, atau proyek yang sesuai dengan tujuan pembelajaran.

ANATES adalah perangkat lunak yang digunakan untuk menganalisis kualitas soal ujian, khususnya dalam hal validitas, reliabilitas, daya pembeda, dan tingkat kesulitan. ANATES dapat membantu pendidik dalam mengevaluasi dan memperbaiki soal ujian sehingga lebih efektif dalam mengukur kemampuan siswa. Beberapa fitur utama ANATES yang relevan dengan penelitian ini adalah:

- **Analisis Butir Soal:** Mengukur validitas dan reliabilitas soal berdasarkan data hasil ujian.
- **Analisis Daya Pembeda:** Mengidentifikasi seberapa baik soal dapat membedakan antara siswa yang memiliki kemampuan tinggi dan rendah.
- **Analisis Kesulitan Soal:** Mengukur seberapa mudah atau sulit soal bagi sebagian besar siswa.
- **Item Response Theory (IRT):** Digunakan untuk menganalisis hubungan antara kemampuan siswa dan kemungkinan mereka menjawab soal dengan benar.

Mata pelajaran informatika di SMK bertujuan untuk memberikan pengetahuan dan keterampilan terkait dengan teknologi informasi, seperti pemrograman, jaringan komputer, dan pengelolaan data. Evaluasi dalam mata pelajaran informatika tidak hanya berfokus pada aspek teori, tetapi juga pada penerapan praktik yang relevan dengan perkembangan teknologi informasi yang cepat. Oleh karena itu, kualitas soal evaluasi harus memperhatikan aspek teori dan praktik informatika yang sesuai dengan standar kompetensi.

Analisis kualitas soal ujian penting untuk memastikan bahwa evaluasi yang dilakukan benar-benar mencerminkan kemampuan siswa dalam hal yang relevan dengan tujuan pembelajaran. Dalam hal ini, kualitas soal yang dievaluasi dengan menggunakan ANATES dapat meningkatkan efektivitas evaluasi pembelajaran di SMK, serta memberikan umpan balik yang berguna bagi pengembangan kurikulum dan metode pengajaran yang lebih baik.

Teori pengukuran pendidikan, khususnya teori klasik dan teori respons butir (Item Response Theory, IRT), dapat digunakan untuk memahami cara mengukur kemampuan siswa melalui soal ujian. Dalam teori klasik, reliabilitas dan validitas soal diuji berdasarkan konsistensi dan kesesuaian soal dengan tujuan evaluasi. Sedangkan dalam

IRT, diperkenalkan konsep seperti tingkat kesulitan soal dan kemampuan siswa yang dapat dianalisis lebih mendalam menggunakan perangkat lunak seperti ANATES.

Penelitian sebelumnya terkait dengan penggunaan ANATES dalam analisis kualitas soal ujian di berbagai disiplin ilmu menunjukkan bahwa perangkat ini sangat berguna dalam meningkatkan kualitas evaluasi pembelajaran. Misalnya, penelitian yang dilakukan di sekolah-sekolah dengan mata pelajaran matematika, bahasa Indonesia, dan ilmu sosial menunjukkan bahwa ANATES dapat memberikan analisis yang mendalam dan memberikan umpan balik yang berguna untuk perbaikan soal evaluasi.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian deskriptif. Tujuan penelitian adalah untuk mengevaluasi kualitas soal ujian mata pelajaran Informatika pada siswa kelas XII AKL 4 SMKN 1 Ngawi. Evaluasi ini dilakukan melalui analisis statistik menggunakan perangkat lunak ANATES untuk menilai aspek reliabilitas, daya pembeda, tingkat kesukaran, dan validitas soal.

Data penelitian diperoleh dari dokumen soal ujian beserta hasil jawaban siswa pada ujian yang telah dilaksanakan. Populasi penelitian adalah seluruh siswa kelas XII AKL 4, dan data diambil secara keseluruhan (total sampling) untuk memastikan representasi penuh. Soal ujian yang dianalisis berjumlah 15 butir pilihan ganda.

Analisis data dilakukan secara bertahap. Pertama, reliabilitas soal diukur untuk menilai konsistensi tes dalam mengukur kemampuan siswa. Kedua, daya pembeda dianalisis untuk menentukan sejauh mana soal mampu membedakan siswa berkemampuan tinggi dan rendah. Ketiga, tingkat kesukaran soal dievaluasi untuk mengelompokkan soal ke dalam kategori mudah, sedang, atau sulit. Keempat, validitas soal ditinjau untuk memastikan kesesuaian setiap butir soal dengan tujuan pembelajaran yang diukur.

Hasil analisis disajikan dalam bentuk table untuk memberikan gambaran yang jelas mengenai kualitas soal ujian. Temuan ini kemudian diinterpretasikan untuk mengidentifikasi kekuatan dan kelemahan soal serta memberikan rekomendasi perbaikan. Pendekatan ini bertujuan untuk memberikan evaluasi yang obyektif dan komprehensif terhadap instrumen evaluasi yang digunakan.

HASIL DAN PEMBAHASAN

Berpikir komputasional merupakan pendekatan sistematis dalam menyelesaikan masalah yang menjadi dasar pembelajaran Informatika. Elemen utamanya meliputi:

1. Dekomposisi

Pemecahan masalah menjadi bagian-bagian kecil yang lebih mudah dikelola, seperti memecah pengembangan aplikasi menjadi desain antarmuka dan logika pemrograman.

2. Pengenalan Pola

Mengidentifikasi pola atau kesamaan dalam data untuk membantu analisis dan prediksi solusi. Contohnya, pengenalan pola pada data siswa untuk membuat algoritma prediktif.

3. Abstraksi

Menyederhanakan masalah dengan hanya fokus pada informasi yang relevan, seperti menggunakan diagram alur untuk merancang algoritma.

4. Perancangan Algoritma

Merancang langkah-langkah sistematis untuk menyelesaikan masalah, misalnya algoritma pengurutan data.

5. Debugging dan Iterasi

Memeriksa kesalahan dalam algoritma atau program dan memperbaiki melalui proses berulang untuk solusi optimal.

Elemen-elemen ini memungkinkan siswa memahami konsep kompleks dan menyelesaikan masalah berbasis teknologi. Implementasinya juga mendukung pengembangan keterampilan analitis dan pemecahan masalah yang penting untuk literasi digital di era modern.

Analisis Berdasarkan Aplikasi ANATES

1. Reliabilitas Tes

Reliabilitas tes adalah ukuran konsistensi suatu instrumen dalam mengukur kemampuan atau atribut tertentu. Hasil analisis menunjukkan bahwa reliabilitas tes sebesar 0,14, yang tergolong sangat rendah berdasarkan kriteria umum dalam evaluasi instrumen (misalnya, reliabilitas $\geq 0,70$ dianggap baik). Nilai ini mengindikasikan bahwa instrumen tes kurang konsisten dalam mengukur

kemampuan siswa. Beberapa faktor yang dapat menjelaskan rendahnya reliabilitas ini antara lain:

a. **Jumlah butir soal yang sedikit**

Reliabilitas instrumen tes cenderung meningkat seiring dengan bertambahnya jumlah butir soal, karena semakin banyak item, semakin besar pula kemungkinan instrumen dapat mencakup berbagai aspek kemampuan yang diukur. Dalam hal ini, jumlah soal yang terbatas mungkin tidak cukup untuk menangkap keseluruhan kemampuan siswa secara representatif, sehingga menurunkan konsistensinya.

b. **Soal yang kurang representatif terhadap tujuan pembelajaran**

Soal yang tidak sesuai atau kurang mencerminkan indikator dan tujuan pembelajaran yang diharapkan dapat menyebabkan siswa menjawab berdasarkan faktor lain selain kemampuan yang hendak diukur. Akibatnya, skor siswa menjadi tidak konsisten dan reliabilitas instrumen menurun.

c. **Variasi tingkat kesukaran soal**

Jika terdapat soal yang terlalu sulit atau terlalu mudah, maka daya diskriminasi antar siswa menjadi kurang efektif. Hal ini berkontribusi pada rendahnya kemampuan instrumen untuk mengukur perbedaan kemampuan siswa secara andal.

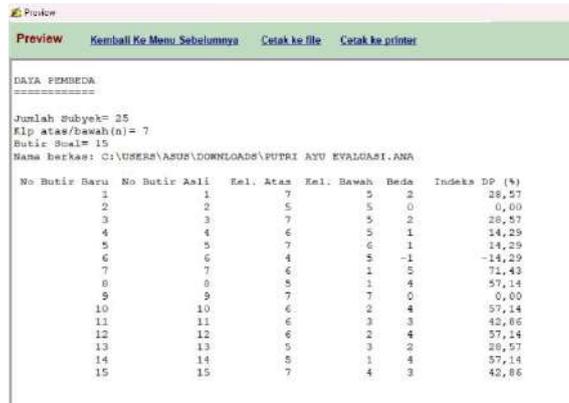
d. **Kualitas penyusunan soal**

Instrumen tes yang memiliki soal ambigu atau tidak jelas dapat mengganggu interpretasi siswa, sehingga menurunkan reliabilitas. Hal ini menunjukkan pentingnya uji coba soal sebelum digunakan secara luas.

No. Urut	No.	Subjek	Kode/Nama Subjek	Skor Ganjil	Skor Genap	Skor Total
1	1	RISSA JNE TABRATU		0	7	15
2	14	FUOI LESDARI		7	6	13
3	15	FUOHU NOVITASARI		7	6	13
4	9	SEPTA WENDUNG		8	4	12
5	11	SEVI AMINAH		7	5	12
6	12	PRISATIA WUR...		7	5	12
7	17	FUOHU KEKAS E...		8	6	14
8	2	RISSA DHT P...		8	3	11
9	14	FUOI WIRANDA ...		6	5	11
10	19	REBI DEPTIARA		6	5	11
11	23	RIKA BELLA SA...		5	6	11
12	20	REYANITA FITR...		5	5	10
13	21	RIFFY PERASA...		4	6	10
14	22	RIFFY JULIANERA		5	5	10
15	3	RISSA EPRIDI		6	2	8
16	8	RISSA NOLLIANY		7	2	9
17	5	RISSA ANDRIYANI		6	3	9
18	10	RISSA DWI CH...		4	5	9
19	24	RISSA PANGKAWATI		4	5	9
20	5	ROSY LAURA SUPRI		6	2	8
21	4	SELI DIAN NOV...		5	3	8
22	10	SYMPA RISDI		5	3	8
23	13	SOPHIA KHUMBA		5	3	8
24	7	RISSA ALDISSA		5	2	7
25	25	RISSA NUCI SA...		4	3	7

Gambar 1. Hasil Reliabilitas Tes

2. Daya Pembeda



No Butir Baru	No Butir Asli	Hel. Atas	Hel. Bawah	Beda	Indeks DP (%)
1	1	7	5	2	28,57
2	2	5	5	0	0,00
3	3	7	5	2	28,57
4	4	6	5	-1	-14,29
5	5	7	6	1	14,29
6	6	4	5	-1	-14,29
7	7	6	1	5	71,43
8	8	5	1	4	57,14
9	9	7	7	0	0,00
10	10	6	2	4	57,14
11	11	6	3	3	42,86
12	12	6	2	4	57,14
13	13	5	3	2	28,57
14	14	5	1	4	57,14
15	15	7	4	3	42,86

Gambar 2. Daya Pembeda

1. Daya Pembeda Positif:

Jika indeks daya pembeda (DP) positif, berarti soal tersebut mampu membedakan antara peserta dengan kemampuan tinggi (kelompok atas) dan rendah (kelompok bawah).

Contoh: Butir soal 1 memiliki indeks DP sebesar **28,57%**, menunjukkan soal cukup baik dalam membedakan kelompok atas dan bawah.

2. Daya Pembeda Negatif:

Jika nilai DP negatif (contoh pada soal 6: **-14,29%**), berarti peserta dengan kemampuan rendah cenderung menjawab benar lebih banyak daripada peserta dengan kemampuan tinggi. Ini menunjukkan bahwa soal tersebut mungkin bermasalah.

3. Kategori Daya Pembeda:

DP < 0%: Soal sangat buruk.

DP 0%–20%: Soal kurang baik.

DP 21%–40%: Soal cukup baik.

DP 41%–60%: Soal baik.

DP > 60%: Soal sangat baik.

Berdasarkan data, beberapa soal (misalnya, soal 4 dan 6) memiliki daya pembeda negatif atau sangat rendah.

4. Faktor Penyebab Daya Pembeda Rendah

a. Konstruksi Soal Tidak Jelas

Soal yang tidak sesuai dengan kompetensi yang diukur atau kalimatnya ambigu menjadi salah satu penyebab utama rendahnya validitas dan daya pembeda soal. Beberapa hal yang mungkin terjadi adalah:

- **Ketidaksesuaian dengan Kompetensi Dasar:** Soal tidak merefleksikan indikator pembelajaran yang diharapkan. Contohnya, jika soal bertujuan untuk mengukur kemampuan berpikir kritis, namun justru hanya menguji hafalan fakta, maka soal tersebut tidak relevan.
- **Ambiguitas Kalimat:** Soal yang dirumuskan dengan bahasa yang tidak jelas atau memiliki makna ganda dapat membingungkan peserta. Akibatnya, siswa yang sebenarnya memiliki kemampuan tinggi pun dapat salah menjawab karena kesalahpahaman.
- **Tidak Ada Konteks yang Jelas:** Soal yang terlalu abstrak tanpa memberikan konteks atau situasi nyata juga menyulitkan siswa dalam memahami maksud soal, sehingga mengurangi efektivitas instrumen. Solusi: Perbaikan konstruk soal dapat dilakukan melalui uji coba awal (pilot testing), pemberian pelatihan penyusunan soal bagi guru, serta melibatkan tim ahli untuk melakukan validasi soal sebelum digunakan.

b. Distraktor Tidak Berfungsi

Distraktor (pilihan jawaban yang salah) memiliki fungsi untuk mengalihkan peserta yang tidak memiliki pemahaman mendalam terhadap materi. Jika distraktor tidak efektif, maka daya pembeda soal akan menurun. Beberapa masalah yang sering terjadi:

- **Distraktor Terlalu Mudah Ditebak:** Pilihan jawaban salah terlalu jauh dari jawaban yang benar atau tidak logis, sehingga peserta dapat menebaknya dengan mudah meskipun tidak memahami materi.
- **Distraktor Tidak Relevan:** Pilihan jawaban yang tidak sesuai dengan konteks soal atau terlalu seragam sehingga tidak memberikan tantangan kepada peserta.
- **Tidak Menarik Perhatian:** Distraktor yang tidak sesuai dengan kesalahan umum yang sering dilakukan siswa tidak akan menarik perhatian mereka, sehingga jawaban benar menjadi lebih mudah ditebak.

Solusi: Pembuatan distraktor perlu berdasarkan analisis kesalahan umum siswa dalam memahami materi, serta memastikan setiap pilihan jawaban memiliki peluang untuk dipilih.

c. Tingkat Kesulitan Tidak Tepat

Tingkat kesulitan soal yang tidak sesuai, baik terlalu mudah maupun terlalu sulit, berdampak pada kemampuan soal untuk membedakan siswa dengan tingkat kemampuan yang berbeda.

- Soal Terlalu Mudah: Soal yang dapat dijawab oleh hampir semua siswa tidak memberikan informasi tentang kemampuan siswa, karena tidak ada variasi skor yang signifikan.
- Soal Terlalu Sulit: Soal yang hanya dapat dijawab oleh sedikit siswa membuatnya gagal mengukur kemampuan mayoritas peserta. Hal ini sering terjadi pada soal yang tidak sesuai dengan tingkat perkembangan atau kemampuan siswa.
Solusi: Menyusun soal dengan distribusi tingkat kesulitan yang seimbang (mudah, sedang, sulit) sesuai prinsip tes yang baik, sehingga dapat mencakup spektrum kemampuan siswa secara keseluruhan.

d. Kesalahan Kunci Jawaban

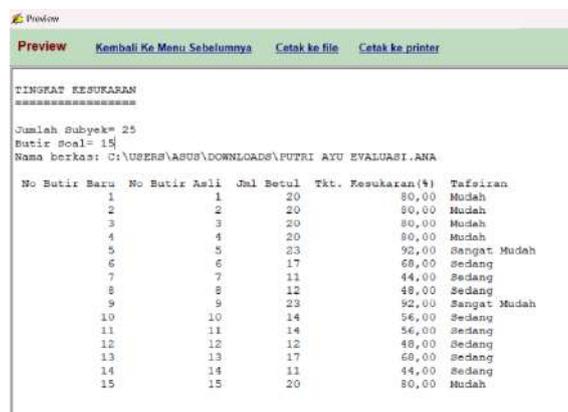
Kunci jawaban yang salah dapat menyebabkan peserta dengan kemampuan tinggi memilih jawaban yang salah, sementara peserta dengan kemampuan rendah memilih jawaban yang benar. Akibatnya, validitas dan daya pembeda soal menjadi terganggu.

- Kesalahan Input: Kunci jawaban salah karena kesalahan teknis dalam memasukkan data atau hasil pengetikan.
- Kurangnya Verifikasi: Kunci jawaban tidak diperiksa ulang oleh penyusun soal atau pihak lain yang kompeten.
- Perubahan Soal Tanpa Koreksi Kunci Jawaban: Ketika soal diubah atau direvisi, namun kunci jawaban tidak diperbarui sesuai perubahan tersebut.
Solusi: Melakukan uji validitas internal soal, termasuk memeriksa kunci jawaban melalui pengoreksian silang oleh tim ahli sebelum distribusi soal.

e. Tingkat Kesukaran

Tingkat kesukaran soal adalah persentase peserta yang menjawab soal dengan benar. Semakin tinggi persentasenya, semakin mudah soal tersebut. Biasanya, interpretasi tingkat kesukaran diklasifikasikan sebagai berikut:

- **80%–100%:** Soal sangat mudah.
- **60%–79%:** Soal mudah.
- **40%–59%:** Soal sedang.
- **20%–39%:** Soal sulit.
- **0%–19%:** Soal sangat sulit.



No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)	Tafsiran
1	1	20	80,00	Mudah
2	2	20	80,00	Mudah
3	3	20	80,00	Mudah
4	4	20	80,00	Mudah
5	5	23	92,00	Sangat Mudah
6	6	17	68,00	Sedang
7	7	11	44,00	Sedang
8	8	12	48,00	Sedang
9	9	23	92,00	Sangat Mudah
10	10	14	56,00	Sedang
11	11	14	56,00	Sedang
12	12	12	48,00	Sedang
13	13	17	68,00	Sedang
14	14	11	44,00	Sedang
15	15	20	80,00	Mudah

Gambar 3. Tingkat Kesukaran

Hasil analisis menunjukkan bahwa mayoritas soal berada dalam kategori mudah hingga sangat mudah, misalnya:

- a. Soal 1, 2, 3, 4, 5, 8, 10, 12, 15 memiliki tingkat kesukaran 80%, termasuk kategori mudah.
- b. Soal 6, 7, 9, 11, 13, 14 memiliki tingkat kesukaran 60%–66,67%, termasuk kategori sedang.

Tingkat kesukaran soal idealnya tersebar dari kategori sulit hingga mudah untuk mencakup seluruh spektrum kemampuan peserta didik. Dalam data ini, tidak ditemukan soal yang masuk kategori sulit atau sangat sulit, yang menunjukkan kurangnya variasi kesukaran. Soal dengan tingkat kesukaran tinggi (80% ke atas) cenderung tidak memberikan tantangan berarti bagi peserta, sehingga kurang efektif dalam mengevaluasi

kemampuan. Soal kategori sedang (40%–60%) adalah yang paling ideal karena mampu membedakan peserta dengan kemampuan tinggi dan rendah.

5. Validitas

Validitas butir soal menunjukkan sejauh mana sebuah soal mampu mengukur apa yang seharusnya diukur secara konsisten. Dalam konteks ini, validitas ditentukan melalui korelasi antara skor pada masing-masing butir dengan skor total peserta didik.

Kriteria Penentuan Validitas :

Nilai korelasi (r) yang signifikan ditentukan berdasarkan batas signifikansi ($p = 0,05$ atau $p = 0,01$) yang disesuaikan dengan derajat kebebasan ($df = N - 2$).

- Pada gambar, dengan **jumlah subyek = 25**, maka $df = 23$. Untuk $p = 0,05$, batas korelasi signifikan adalah $r \geq 0,396$. Untuk $p = 0,01$, batas korelasi signifikan adalah $r \geq 0,505$.

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0,338	-
2	2	-0,010	-
3	3	0,238	-
4	4	0,188	-
5	5	0,087	-
6	6	0,012	-
7	7	0,570	Signifikan
8	8	0,441	-
9	9	0,170	-
10	10	0,551	Signifikan
11	11	0,271	-
12	12	0,480	-
13	13	0,267	-
14	14	0,570	Signifikan
15	15	0,338	-

df (N-2)	PT(0,05)	PT(0,01)	df (N-2)	PT(0,05)	PT(0,01)
10	0,576	0,709	50	0,250	0,325
15	0,482	0,606	70	0,233	0,302

Gambar 4. Validitas Butir Soal

Korelasi antara skor butir dan skor total menunjukkan seberapa baik suatu soal dapat mengukur keseluruhan kemampuan peserta. Nilai korelasi (r) dapat diartikan sebagai berikut:

- **Korelasi** **Positif** **Signifikan:**
Jika nilai korelasi tinggi (misalnya $r > 0,396$ untuk $p < 0,05$), soal tersebut dapat dikatakan valid, artinya butir soal tersebut sejalan dengan kemampuan keseluruhan peserta.

- **Korelasi Rendah atau Tidak Signifikan:** Korelasi yang rendah atau negatif menunjukkan bahwa soal tersebut tidak sejalan dengan kemampuan keseluruhan peserta. Hal ini dapat menunjukkan bahwa soal:
 - Tidak jelas atau ambigu.
 - Tidak relevan dengan materi yang diujikan.
 - Tidak sesuai dengan tingkat kemampuan peserta.

Hasil Validitas Berdasarkan Butir Soal

- **Soal Valid Signifikan:**

Soal-soal berikut memiliki nilai korelasi signifikan (valid) berdasarkan tabel:

- Butir 7 ($r = 0,570$, signifikan pada $p < 0,05$). Butir ini memiliki korelasi cukup tinggi, menunjukkan bahwa soal ini sejalan dengan keseluruhan kemampuan peserta.
- Butir 10 ($r = 0,551$, signifikan pada $p < 0,05$). Korelasi ini menunjukkan bahwa soal cukup kuat dalam mengukur keseluruhan kemampuan peserta.
- Butir 14 ($r = 0,570$, signifikan pada $p < 0,05$). Validitas yang baik menunjukkan bahwa soal ini efektif digunakan.

- **Soal Tidak Valid:**

Soal-soal berikut memiliki nilai korelasi rendah atau negatif, sehingga dianggap tidak valid:

- Butir 2 ($r = -0,010$): Korelasi negatif menunjukkan bahwa peserta yang memiliki skor total tinggi justru cenderung salah menjawab soal ini, sehingga soal perlu diperbaiki.
- Butir 5 ($r = 0,097$): Korelasi rendah menunjukkan bahwa soal ini kurang mampu mengukur kemampuan peserta secara konsisten.
- Soal lain dengan nilai korelasi di bawah **0,396** tidak menunjukkan hubungan yang signifikan dengan skor total.

Berdasarkan hasil validitas butir soal di dapat soal tes ini belum optimal, karena mayoritas soal tidak memiliki korelasi signifikan dengan skor total. Hanya beberapa soal saja yang valid signifikan, sehingga tes ini memerlukan perbaikan substansial untuk meningkatkan kualitas dan validitasnya.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis soal didapat Kesimpulan diantaranya sebagai berikut :

1. Kualitas Instrumen Evaluasi Tes

Berdasarkan hasil analisis, instrumen tes menunjukkan kelemahan signifikan pada aspek reliabilitas, daya pembeda, tingkat kesukaran, dan validitas soal. Reliabilitas tes yang sangat rendah (**0,14**) menunjukkan kurangnya konsistensi dalam mengukur kemampuan siswa. Sebagian besar soal memiliki daya pembeda yang lemah atau negatif, serta tingkat kesukaran yang didominasi oleh kategori **mudah** hingga **sedang**, sehingga kurang mampu mencakup spektrum kemampuan siswa secara keseluruhan. Mayoritas soal juga tidak memenuhi kriteria validitas signifikan, menandakan perlunya revisi mendalam pada butir soal.

2. Kontekstualisasi Dalam Pembelajaran Informatika

Tes ini bertujuan untuk mengukur kemampuan siswa dalam berpikir komputasional, yang mencakup elemen penting seperti dekomposisi, pengenalan pola, abstraksi, perancangan algoritma, debugging, dan iterasi. Namun, kelemahan pada tes ini dapat menghambat upaya untuk mengevaluasi kompetensi siswa secara akurat. Instrumen yang tidak memadai akan sulit mendukung literasi digital dan keterampilan berpikir kritis yang menjadi target utama pembelajaran informatika.

3. Peluang Perbaikan dan Penyempurnaan Tes

Analisis menunjukkan bahwa instrumen ini membutuhkan perbaikan menyeluruh, mulai dari penambahan jumlah soal untuk meningkatkan reliabilitas, hingga revisi kualitas soal agar lebih sesuai dengan tujuan pembelajaran. Distribusi tingkat kesukaran yang lebih seimbang (mudah, sedang, sulit) perlu diperhatikan agar tes mampu menilai kemampuan siswa dengan lebih representatif. Daya pembeda soal juga perlu ditingkatkan agar lebih efektif dalam membedakan siswa berdasarkan tingkat kemampuannya.

4. Dampak Positif Perbaikan Instrumen

Dengan penyempurnaan instrumen evaluasi, guru dan pendidik akan memiliki alat ukur yang lebih valid dan andal untuk mengevaluasi keberhasilan pembelajaran. Hal ini akan membantu siswa memahami konsep informatika secara lebih mendalam, memperkuat kemampuan berpikir komputasional, dan mendukung pengembangan keterampilan abad ke-21, seperti pemecahan masalah, logika analitis, dan kreativitas.

Instrumen yang baik juga memungkinkan pengambilan keputusan berbasis data dalam perencanaan pembelajaran.

5. Perbaikan yang bisa dilakukan:

- **Pengembangan Instrumen:** Lakukan uji coba soal pada kelompok yang lebih besar, analisis butir secara berulang, dan revisi berdasarkan hasil analisis statistik. Pastikan soal mencerminkan indikator pembelajaran yang diharapkan.
- **Pelatihan Penyusunan Soal:** Pelatihan bagi penyusun soal dapat meningkatkan pemahaman tentang desain soal yang valid, andal, dan sesuai dengan karakteristik siswa.
- **Evaluasi Berbasis Teknologi:** Memanfaatkan platform digital untuk analisis data hasil tes secara otomatis agar mempermudah identifikasi kelemahan soal dan pengukuran kemampuan siswa.

Secara keseluruhan, instrumen tes saat ini belum optimal dalam mengukur kemampuan siswa dalam berpikir komputasional. Penyempurnaan yang terarah akan meningkatkan kualitas evaluasi dan memberikan dampak positif pada proses pembelajaran informatika, mendukung tujuan pendidikan yang berfokus pada pengembangan keterampilan berpikir tingkat tinggi, literasi digital, dan adaptasi terhadap tantangan era modern.

DAFTAR REFERENSI

- Abdillah, L., & Zulkifli. (2020). Evaluasi Validitas dan Reliabilitas Butir Soal Pendidikan. *Jurnal Pendidikan*, 12(3), 45-56.
- Arikunto, S. (2016). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Arikunto, S. (2019). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Budiyono, S. (2017). Analisis reliabilitas tes dengan pendekatan klasik. *Jurnal Penelitian dan Evaluasi Pendidikan*, 21(1), 15–28.
- Fitriani, R. (2021). Analisis Validitas Instrumen Evaluasi Berbasis Komputer. *Jurnal Teknologi Pendidikan*, 10(2), 234-245.

- Grover, S., & Pea, R. (2018). Computational thinking in K–12: A review of the state of the field. *Educational Researcher*, 47(1), 16-24.
- Lye, S. Y., & Koh, J. H. L. (2018). Review on teaching and learning of computational thinking through programming. *Computers in Human Behavior*, 41, 51-61.
- Nugraha, R. (2021). Teknik Analisis Butir Soal Menggunakan Software. *Jurnal Teknologi Pendidikan*, 15(1), 45-58.
- Nurgiyantoro, B., Gunawan, & Marzuki. (2015). *Statistik Terapan untuk Penelitian Ilmu-Ilmu Sosial*. Yogyakarta: Gadjah Mada University Press.
- Permendikbud No. 37 Tahun 2018. Tentang Kompetensi Inti dan Kompetensi Dasar Pelajaran pada Kurikulum 2013.
- Setiawan, A. (2020). Pengembangan Instrumen Evaluasi Pendidikan. Surabaya: Universitas Negeri Surabaya Press.
- Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142-158.
- Suherman, A. (2019). Peningkatan kualitas instrumen evaluasi pendidikan. *Jurnal Pendidikan dan Kebudayaan*, 23(3), 251–262.
- Supriyanto, A., & Suyadi. (2023). Analisis Statistik dalam Evaluasi Pendidikan. Bandung: Alfabeta.
- Susanto, H., & Marlina, N. (2022). Analisis Daya Pembeda sebagai Instrumen Evaluasi. *Journal of Educational Studies*, 14(2), 123-135.
- Susanto, H., & Marlina, N. (2022). Kualitas Butir Soal: Pendekatan Validitas dan Reliabilitas. *Journal of Educational Studies*, 14(2), 123-135.
- Wibowo, A. (2020). Evaluasi kualitas tes berbasis uji coba lapangan. *Jurnal Evaluasi Pendidikan*, 10(2), 89–97.
- Wing, J. M. (2017). Computational thinking's influence on research and education for all. *Italian Journal of Educational Technology*, 25(2), 7-14.