

ANALISIS BUTIR SOAL KEMAMPUAN BERPIKIR KRITIS PADA MATERI TEKS DESKRIPTIF DALAM PEMBELAJARAN BAHASA INGGRIS KELAS 7 DI SMP NEGERI 2 RANGKASBITUNG

Sumaryani Rita Susilowati, Isti Rusdiyani, Aan Hendrayana
Universitas Sultan Ageng Tirtayasa, Banten, Indonesia
Corresponding author email: sumaryanis70@guru.smp.belajar.id

Abstrak:

Penelitian ini dilakukan dengan tujuan merancang serta mengembangkan instrumen evaluasi yang mampu mengukur keterampilan berpikir kritis pada materi teks deskriptif dalam pembelajaran Bahasa Inggris. Fokus utamanya adalah memastikan validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan kesesuaian instrumen dengan tujuan pembelajaran, sehingga dapat mendukung siswa dalam berpikir kritis secara efektif dan terstruktur. Penelitian ini termasuk dalam jenis penelitian deskriptif yang bertujuan untuk memperoleh data terkait tingkat kesulitan, kemampuan diskriminasi, reliabilitas, serta validitas setiap butir soal dalam menilai keterampilan berpikir kritis pada materi teks deskriptif. Dari 25 butir tes yang diuji, sebanyak 18 butir (72%) dinyatakan valid, sedangkan 7 butir lainnya tidak valid karena tidak memenuhi salah satu atau kedua kriteria yang ditetapkan. Temuan ini menunjukkan bahwa sebagian besar butir tes telah mampu mengukur secara akurat, sementara butir yang tidak valid memerlukan revisi untuk perbaikan sebelum digunakan.

Kata kunci: Instrumen Penilaian, Kemampuan Berpikir Kritis, Teks Deskriptif

Abstract:

This research seeks to design and develop an evaluation tool tailored to assess critical thinking skills on descriptive text material in English learning. The primary focus is on ensuring the accuracy, consistency, level of difficulty, distinguishing capability, and alignment of the instrument with learning objectives to support students in thinking critically in an effective and structured manner. This research falls under the descriptive study category, aiming to obtain information about the difficulty level, discrimination power, reliability, and validity of each test item in measuring critical thinking skills on descriptive text material. Out of the 25 test items examined, 18 items (72%) were found to be valid, while 7 items were deemed invalid for failing to meet one or both of the established criteria. These findings indicate that the majority of test items were able to measure accurately, whereas the invalid items require revisions for improvement before being utilized.

Keywords: Assessment Instrument, Critical Thinking Skills, Descriptive Text

Pendahuluan

Kompetensi utama yang menjadi kebutuhan utama di era abad ke-21 meliputi empat keterampilan penting: (1) berpikir kritis (critical thinking), yaitu kemampuan siswa dalam menyelesaikan masalah secara kontekstual dengan menggunakan logika yang kritis dan rasional; (2) kreativitas (creativity), yakni kapasitas untuk membantu siswa menghasilkan berbagai solusi yang inovatif; (3) kolaborasi (collaboration), mendukung siswa untuk bekerja sama secara kelompok, bersikap toleran, dan berkolaborasi untuk meraih tujuan bersama; serta (4) komunikasi (communication), yang membantu peserta didik menyampaikan gagasan secara efektif dan memahami informasi dengan baik (Wilcox, D., Liu, J. C., Thall, J., & Howley, 2017). Sebagai upaya mengimplementasikan keempat kompetensi tersebut dalam menghadapi tantangan abad ke-21, kebijakan pendidikan di Indonesia mengadopsi konsep *Higher Order Thinking Skills* (HOTS). HOTS mengacu pada keterampilan kognitif tingkat tinggi, seperti kemampuan menganalisis dan mengevaluasi, yang dapat diajarkan oleh guru kepada peserta didik (Aulia, 2021).

Dalam konteks pembelajaran bahasa, kompetensi ini berperan penting karena pembelajaran bahasa melibatkan pengembangan empat keterampilan utama, yaitu mendengarkan, berbicara, membaca, dan

menulis (Batubara, 2021). Dengan demikian, pengajaran bahasa tidak berfokus pada aspek kebahasaan tetapi pada peningkatan keterampilan abad ke-21, terutama berpikir kritis. Kemampuan berpikir kritis mencakup aktivitas analisis ide secara mendalam, membedakan dengan cermat, mengevaluasi, dan mengembangkan gagasan. Keterampilan ini menjadi elemen penting dalam pembelajaran abad ke-21, terutama karena didukung oleh teknologi yang mempermudah akses, analisis, dan penyampaian informasi (Rahayu et al., 2019).

Menurut (Saputra, 2020), berpikir kritis adalah proses terstruktur yang melibatkan aktivitas seperti merumuskan masalah, memberikan argumen, melakukan evaluasi, hingga mengambil keputusan. Kemampuan berpikir kritis sangat relevan dalam pembelajaran bahasa Inggris, terutama pada materi teks deskriptif. Pada jenis teks ini, siswa dituntut untuk mendeskripsikan orang, tempat, atau benda secara rinci, sambil mempraktikkan keterampilan berpikir kritis, seperti menganalisis struktur teks, mengevaluasi penggunaan kata sifat, dan mengembangkan deskripsi logis.

Untuk mengevaluasi keterampilan berpikir kritis siswa dalam pembelajaran bahasa Inggris, diperlukan instrumen penilaian yang dirancang khusus. Berdasarkan kerangka kerja Facione (1990), kemampuan berpikir kritis melibatkan enam keterampilan utama: interpretasi, analisis, evaluasi, inferensi, eksplanasi, dan refleksi. Dalam konteks teks deskriptif, instrumen penilaian dapat mencakup analisis penggunaan kosakata, evaluasi struktur teks, dan pengembangan ide yang relevan (Zahro et al., 2024). Namun, penelitian menunjukkan bahwa kemampuan guru dalam merancang tes untuk menilai kemampuan berpikir kritis masih terbatas (Laili, M., Aini, N., & Christanti, 2020). Oleh karena itu, diperlukan pelatihan dan penelitian lebih lanjut untuk mengembangkan instrumen yang tidak hanya mengevaluasi pencapaian pembelajaran, namun juga mendorong pembelajaran yang bermakna.

Penelitian ini bertujuan untuk mengembangkan instrumen evaluasi yang dirancang khusus untuk menilai keterampilan berpikir kritis dalam materi teks deskriptif dalam pembelajaran bahasa Inggris. Fokus utama adalah pada validitas, reliabilitas, tingkat kesukaran, daya pembeda dan kesesuaian instrumen dengan tujuan pembelajaran, sehingga dapat mendukung siswa untuk berpikir kritis secara efektif dan terstruktur.

Metode Penelitian

Penelitian ini merupakan jenis penelitian deskriptif yang bertujuan untuk memperoleh informasi terkait tingkat kesulitan, kemampuan setiap butir soal dalam membedakan tingkat kemampuan, serta reliabilitas dan validitasnya dalam mengukur keterampilan berpikir kritis pada materi teks deskriptif. Data dianalisis menggunakan metode teknik kuantitatif, meliputi uji validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Butir soal yang dirancang diujikan pada 30 responden pada bulan November, kelas 7 di SMP Negeri 2 Rangkasbitung. Jumlah soal yang diujikan sebanyak 25 butir soal pilihan ganda.

Hasil Penelitian dan Pembahasan

Hasil validasi oleh para ahli materi diperoleh melalui lembar angket yang mencakup penilaian dari validator ahli materi. Data dari angket tersebut dihitung untuk menentukan persentase dari setiap aspek penilaian yang diberikan oleh para ahli. Para ahli juga memberikan saran dan masukan yang menjadi dasar dalam memperbaiki beberapa soal, yaitu penggunaan gambar untuk beberapa soal efektif membantu pemahaman siswa namun untuk meningkatkan kualitas, perlu ditambahkan elemen berpikir kritis.

Berdasarkan hasil penilaian ahli materi, total skor yang diperoleh adalah 78 dari skor maksimum 80, dengan rincian nomor 13 (Memuat kriteria soal HOTS) dan nomor 15 (Gambar dan narasi jelas dan berfungsi) masing-masing mendapatkan skor 3, sementara 18 butir lainnya mendapatkan skor 4. Persentase pencapaian dihitung dengan rumus:

$$\text{Persentase} = \frac{\text{Skor yang diperoleh} \times 100}{\text{Skor maksimum}}$$

Sehingga diperoleh nilai 97,5%, yang menempatkan instrumen soal dalam kategori Sangat Baik (70-80).

Terdapat dua butir soal yang memerlukan perbaikan, yaitu nomor 13 dan 15. Skor 3 pada nomor 13 menunjukkan bahwa soal belum sepenuhnya memenuhi kriteria HOTS karena kurang merangsang kemampuan berpikir tingkat tinggi atau pemecahan masalah kompleks. Skor 3 pada nomor 15 mengindikasikan bahwa gambar atau narasi yang digunakan masih kurang jelas atau tidak berfungsi optimal dalam mendukung pemahaman peserta didik. Secara keseluruhan, sebanyak 18 butir soal lainnya telah memenuhi kriteria penilaian dengan sangat baik dari segi konten, struktur, maupun penggunaan bahasa. Dengan demikian, instrumen soal dinyatakan Sangat Baik dengan total skor 78 dan persentase 97,5%, namun tetap memerlukan perbaikan pada butir soal nomor 13 dan 15 agar lebih optimal.

Revisi Taksonomi Bloom oleh Anderson dan Krathwohl bertujuan untuk menyesuaikannya dengan kebutuhan pendidikan modern, terutama terkait dengan *Higher Order Thinking Skills* (HOTS) pada ranah kognitif. Revisi ini meliputi tiga tingkat kemampuan utama, yaitu analisis (*analyze*), evaluasi (*evaluate*), dan kreasi (*create*). Tingkat kesulitan soal ditentukan dengan cara mengidentifikasi perbedaan jumlah responden yang menjawab benar terhadap total jumlah responden. Hasil perhitungan tersebut kemudian dibandingkan dengan kriteria Indeks Kesulitan (IK). Sementara itu, daya pembedanya setiap soal pilihan ganda merujuk pada kemampuan soal dalam membedakan responden dengan skor yang tinggi dan responden dengan skor yang rendah.

Kemampuan daya pembedanya dianalisis melalui mengelompokkan nilai r-hitungnya yang diperoleh melalui uji Pearson Product Moment untuk setiap butir tes, kemudian dibandingkan dengan Indeks Daya Pembeda (IDP). Validitas soal dievaluasi berdasarkan distribusi nilai yang dihasilkan saat uji coba soal, menggunakan metode Pearson Product Moment. Suatu soal dianggap valid apabila nilai korelasi Pearson atau r-hitung melebihi r-tabel. Validitasnya juga dapat ditentukan dari nilai korelasi Pearson dengan tingkat signifikansi. Jika nilai korelasi Pearson positif dan didukung tingkat signifikansi kurang dari 0,05, soal dinyatakan valid. Sebaliknya, jika tingkat signifikansi terpenuhi (kurang dari 0,05) tetapi korelasi Pearson bernilai negatif, maka soal tetap dianggap tidak valid. Jika tingkat signifikansi lebih dari 0,05, soal juga dianggap tidak valid. Untuk mengukur reliabilitas instrumen, digunakan pedoman tertentu, di mana instrumen dinyatakan reliabel jika nilai Alpha Cronbach lebih besar dari 0,06. Sebaliknya, jika nilai Alpha Cronbach di bawah 0,06, maka instrumen dinyatakan tidak reliabel.

Table 1. Hasil pengujian tingkat kesulitan

Butir soal	Indeks kesukaran	Kriteria
1	0,4	Sedang
2	0,333	Sedang
3	0,433	Sedang
4	0,4	Sedang
5	0,133	Sukar
6	0,667	Sedang
7	0,733	Mudah
8	0,567	Sedang
9	0,767	Mudah
10	0,667	Sedang
11	0,5	Sedang
12	0,467	Sedang
13	0,367	Sedang
14	0,4	Sedang
15	0,467	Sedang
16	0,233	Sukar
17	0,567	Sedang
18	0,533	Sedang
19	0,267	Sukar
20	0,333	Sedang
21	0,533	Sedang
22	0,167	Sukar
23	0,533	Sedang
24	0,333	Sedang
25	0,567	Sedang

Tabel 1 menampilkan hasil pengujian tingkat kesulitan setiap soal. Berdasarkan analisis, butir tes diklasifikasikan ke dalam tiga kategori: sederhana, menengah, dan sulit. Hasil tersebut, 76% butir soal masuk kategori sedang, 16% termasuk kategori sukar, dan 8% tergolong mudah. Visualisasi proporsi setiap kategori butir soal pada Gambar 1.



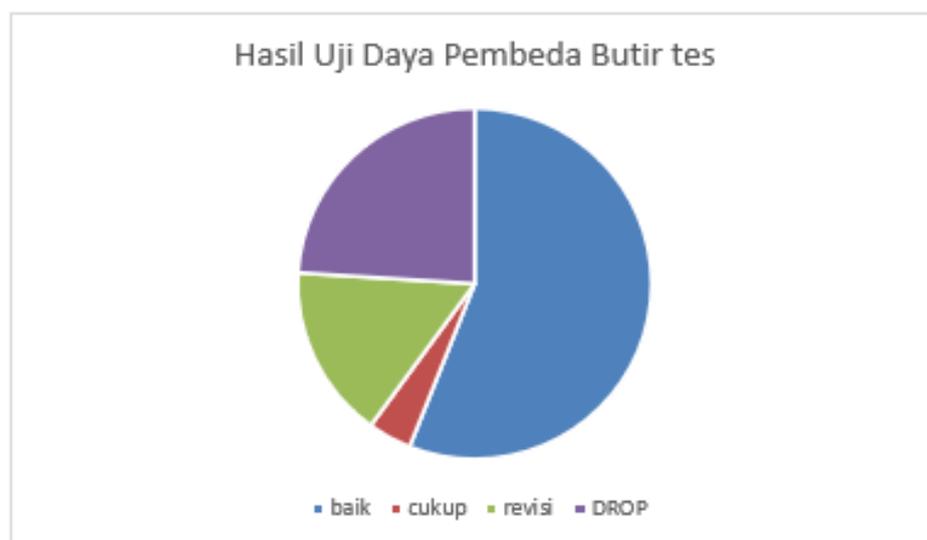
Gambar 1. Klasifikasi Tingkat Kesukaran Butir Tes

Pengujian daya pembeda, yang menunjukkan kemampuan butir tes melalui perbedaan responden dengan kemampuan tinggi atau rendah, menghasilkan empat kategori, seperti baik, cukup, perlu perbaikan, dan rendah atau tidak layak digunakan. Penjelasan lebih detail mengenai hasil pengujian pada Tabel 2.

Table 2 Capaian Uji Daya Pembeda

Butir Soal	IDP	Kriteria
1	0,266	Revisi
2	0,267	Revisi
3	0,333	Cukup
4	0,266	Revisi
5	0	Drop
6	0,533	Baik
7	0,533	Baik
8	0,6	Baik
9	0,467	Baik
10	0,533	Baik
11	0,734	Baik
12	0	Drop
13	-0,067	Drop
14	0,4	Baik
15	0,4	Baik
16	0,2	Revisi
17	0,467	Baik
18	0,533	Baik
19	0	Drop
20	-0,133	Drop
21	0,4	Baik
22	0,067	Drop
23	0,533	Baik
24	0,667	Baik
25	0,6	Baik

Berdasarkan tabel tersebut, persentase butir tes yang tergolong dalam kategori baik 56%, kategori cukup 4%, kategori revisi mencapai 16%, dan kategori drop atau kurang mencakup 24%. Perbandingan persentase keempat kategori ini ditampilkan secara visual pada Gambar 2.



Gambar 2. Hasil Uji Daya Pembeda Butir tes

Pengujian reliabilitas yang digunakan metode Cronbach alpha mendapatkan nilai sebesar 0,772. Berdasarkan hasil ini, dapat disimpulkan bahwa instrumen tersebut memiliki reliabilitas yang baik, karena nilai Cronbach alpha melebihi ambang batas 0,6. Hasil uji reliabilitas ini diperoleh melalui metode Cronbach alpha.

Ada dua metode yang dapat digunakan untuk menentukan validitas suatu butir tes. Metode pertama adalah perbandingan nilai pearson correlation (r_{hitung}) dan r_{tabel} . Jika r_{hitung} lebih besar dari pada r_{tabel} , maka butir tes dianggap valid ($r_{hitung} > r_{tabel}$). Metode kedua melibatkan kombinasi antara nilai pearson correlation dan tingkat signifikansi. Suatu butir tes dinyatakan valid jika nilai pearson correlation bernilai positif dan didukung oleh tingkat signifikansi kurang dari 0,05. Ketika nilai signifikansinya kriteria (kurang dari 0,05) namun nilai korelasi Pearsonnya menunjukkan angka negatif, maka butir tes dianggap tidak valid. Demikian pula, ketika nilai signifikansi melebihi 0,05, butir tes tetap dinyatakan tidak valid.

Hasil dari uji validitas butir tes berdasarkan analisis *Pearson correlation (rhitung)* dan tingkat signifikansi menggunakan metode *Pearson product moment*. Validitas butir tes ditentukan melalui dua kriteria utama, yaitu perbandingan nilai *rhitung* dengan *rtabel* dan kombinasi nilai *rhitung* dengan tingkat signifikansi. Suatu butir tes dianggap diterima jika r_{hitung} melebihi dari r_{tabel} , memiliki nilai yang positif, dan didukung oleh tingkat signifikansi kurang dari 0,05. Sebaliknya, butir soalnya dianggap tidak valid jika *rhitung* bernilai negatif meskipun tingkat signifikansi memenuhi syarat atau jika tingkat signifikansi lebih besar dari 0,05. Dari total 25 butir tes yang diuji, sebanyak 18 butir (72%) dinyatakan valid, sementara 7 butir lainnya tidak valid karena tidak memenuhi salah satu atau kedua kriteria tersebut. Hal ini menunjukkan bahwa mayoritas butir tes telah mampu mengukur secara akurat, sedangkan butir yang tidak valid memerlukan revisi untuk perbaikan sebelum digunakan.

Simpulan

Tes yang dirancang untuk menilai kemampuan berpikir kritis dalam materi teks deskriptif pembelajaran Bahasa Inggris menggunakan format pilihan ganda. Namun, sering kali ditemukan kelemahan pada butir soal yang disusun, sehingga analisis soal diperlukan tes terlebih dahulu. Fokus penelitian ini pada analisis soal yang akan diteskan, seperti aspek kesulitan, daya pembeda soal, reliabilitas soal, dan validitasnya.

Data analisisnya diperoleh melalui pengujian soal tes terhadap 30 siswa. Tingkatan kesulitan dihitung melalui membandingkan jumlah siswa yang menjawab benar dengan total jumlah siswa. Hasil analisisnya bahwa butir soal terbagi ke dalam kategori mudah, sedang, dan sulit. Analisis daya pembedanya dilakukan melalui pengelompokan nilai dari r_{hitung} dari uji Pearson untuk setiap soal sesuai dengan indeks daya pembedanya. Hasilnya, butir soal dikelompokkan ke dalam kategori baik, cukup, memerlukan revisi, dan kurang.

Uji reliabilitas menggunakan metode Cronbach alpha menunjukkan bahwa tes pilihan ganda yang dirancang memiliki tingkat reliabilitas yang memadai. Selain itu, uji validitas menggunakan metode *Pearson product moment* mengungkapkan bahwa 18 dari 25 butir soal, atau 72%, dinyatakan valid.

Referensi

- Aulia, V. (2021). Penyusunan Soal Berbasis Higher Order Thinking Skills Pada Mata Pelajaran Bahasa Inggris. *Prosiding Seminar Nasional Sensasada*, 1(1), 189–197.
- Batubara, S. A. (2021). Kemampuan Berpikir Kritis Siswa Dalam Menulis Teks Deskripsi Menggunakan Media Video Pengetahuan Alam Kelas VII SMP SIMAN JAYA. *Jurnal Bahasa Lingua Scientia*, 13(2), 233–246. <https://doi.org/10.21274/lis.2021.13.2.233-246>
- Facione, P. (2015). Critical Thinking: What It Is and Why It Counts. *Insight Assessment*, 1–30.
- Kemendikbud. (2019). *Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi (Higher Order Thinking Skills) Bahasa Inggris*. Kementerian Pendidikan dan Kebudayaan.
- Laili, M., Aini, N., & Christanti, A. (2020). Higher Order Thinking Skills (HOTS) Dalam Penilaian Bahasa Inggris Siswa SMA. *Lintang Songo: Jurnal Pendidikan*, 3(1), 18–25.
- Nugroho, R. A. (2018). *Higher Order Thinking Skills. Kemampuan Berpikir Tingkat Tinggi. Konsep. Pembelajaran. Penilaian. Penyusunan Soal Sesuai HOTS*. PT Gramedia Widiasarana Indonesia.
- Rahayu, D. M., Roesminingsih, M. V., -, H., & Subroto, W. T. (2019). The Use of Interactive Multimedia to Improve Critical Thinking Skills of Primary School Students. *International Journal of Scientific and Research Publications (IJSRP)*, 9(4), 138–144. <https://doi.org/10.29322/ijsrp.9.04.2019.p8821>
- Saputra, H. (2020). Kemampuan Berfikir Kritis Matematis. *Perpustakaan IAI Agus Salim*. <https://doi.org/10.17605/OSF.IO/TJ76P>
- Wilcox, D., Liu, J. C., Thall, J., & Howley, T. (2017). Integration of Teaching Practice for Students' 21st Century Skills: Faculty Practice and Perception. *International Journal of Technology in Teaching and Learning*, 13(2), 55–77.
- Zahro, S. M., Susanto, & Suwito, A. (2024). Analisis Kemampuan Berpikir Kritis Siswa Kelas Xii Di Jember Pada Materi Dimensi Tiga. *SECONDARY: Jurnal Inovasi Pendidikan Menengah*, 4(2), 55–60. <https://doi.org/10.51878/secondary.v4i2.3004>