

Analisis Soal UTS Fisika Siswa Kelas XI SMA Negeri 1 Batang Kuis Berbasis Comparative Thinking: Validitas, Reliabilitas, Kesukaran, dan Daya Pembeda

Cinta Marcella Namira¹ Jessica Banjarnahor² Lourensia Bravini Lahera³ Ribka Nainggolan⁴ Septi Maria Gultom⁵

Program Studi Pendidikan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Kabupaten Deli Serdang, Provinsi Sumatera Utara, Indonesia^{1,2,3,4,5}

Email: cintamarcella.4233121025@mhs.unimed.ac.id¹

jessicabanjarnahor4233121027@mhs.unimed.ac.id²

lourensia.4233121048@mhs.unimed.ac.id³ ribka.4233121047@mhs.unimed.ac.id⁴

septi.4233121052@mhs.unimed.ac.id⁵

Abstrak

Penelitian ini menganalisis kualitas instrumen evaluasi fisika berbasis comparative thinking melalui parameter validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Metode deskriptif kuantitatif diterapkan pada 20 soal pilihan ganda yang diujikan kepada 36 siswa kelas XI IPA di SMA Negeri Kota Medan. Hasil menunjukkan 4 soal tidak valid (r -hitung $< 0,329$) akibat ketiadaan konteks komparatif, sementara reliabilitas instrumen memenuhi kriteria ($\alpha = 0,749$). Distribusi tingkat kesukaran didominasi soal mudah (50%) dengan indeks $> 0,7$, sedangkan 45% soal memiliki daya pembeda rendah ($D < 0,3$), terutama pada pertanyaan prosedural tanpa analisis multidimensi. Temuan mengungkap ketimpangan antara tuntutan comparative thinking dan struktur soal yang cenderung terisolasi. Penelitian ini merekomendasikan integrasi skenario komparatif berbasis kasus, penyeimbangan parameter psikometrik, serta optimalisasi distraktor untuk meningkatkan validitas konstruk dan daya diagnostik instrumen. Implikasinya, pengembangan soal fisika perlu mengadopsi pendekatan komparatif sistematis guna mengukur kemampuan analitis siswa secara holistik.

Kata Kunci: Comparative Thinking, Validitas, Reliabilitas, Kesukaran Soal, Daya Pembeda, Soal Fisika

Abstract

This study analyzes the quality of physics assessment instruments based on comparative thinking through validity, reliability, difficulty index, and discriminative power parameters. A quantitative descriptive method was applied to 20 multiple-choice questions tested on 36 eleventh-grade science students in a public high school in Medan. Results revealed 4 invalid items (r -count < 0.329) due to the absence of comparative contexts, while instrument reliability met the criteria ($\alpha = 0.749$). The difficulty index distribution was dominated by easy questions (50% with index > 0.7), whereas 45% of items showed low discriminative power ($D < 0.3$), particularly in procedural questions lacking multidimensional analysis. The findings highlight a mismatch between the demands of comparative thinking and the isolated structure of conventional questions. This research recommends integrating case-based comparative scenarios, balancing psychometric parameters, and optimizing distractors to enhance construct validity and diagnostic power. The study implies that physics question development should adopt systematic comparative approaches to holistically assess students' analytical skills.

Keywords: Comparative Thinking, Validity, Reliability, Difficulty Index, Discriminative Power, Physics Questions



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

PENDAHULUAN

Pembelajaran fisika tidak hanya bertujuan untuk mentransfer pengetahuan, tetapi juga untuk mengembangkan keterampilan berpikir kritis dan analitis siswa. Salah satu metode yang dapat meningkatkan pemahaman konseptual siswa adalah Comparative Thinking, yaitu pendekatan yang mengharuskan siswa untuk membandingkan berbagai konsep fisika dan

mengidentifikasi hubungan di antara mereka. Dengan metode ini, siswa tidak hanya menghafal rumus, tetapi juga memahami bagaimana prinsip-prinsip fisika saling berkaitan dan dapat diterapkan dalam berbagai situasi. Konsep validitas dalam asesmen sangat penting dalam memastikan bahwa instrumen evaluasi benar-benar mengukur apa yang seharusnya diukur. Validitas melibatkan berbagai aspek, seperti validitas isi, kriteria, dan konstruk, yang masing-masing berkontribusi dalam menilai kualitas suatu tes (Mehrens & Lehmann, n.d.). Dalam pendidikan fisika, validitas asesmen menjadi krusial untuk mengevaluasi keterampilan berpikir analitis siswa, bukan sekadar hafalan rumus atau prosedur penyelesaian soal. Selain itu, asesmen dalam pendidikan fisika juga harus mempertimbangkan kreativitas saintifik siswa, yang mencakup sikap ilmiah, keterampilan proses, dan pemahaman konsep (Hu & Adey, 2002). Sebuah evaluasi yang efektif tidak hanya mengukur hasil belajar, tetapi juga mendorong siswa untuk mengeksplorasi konsep fisika melalui berpikir kreatif dan inkuiri.

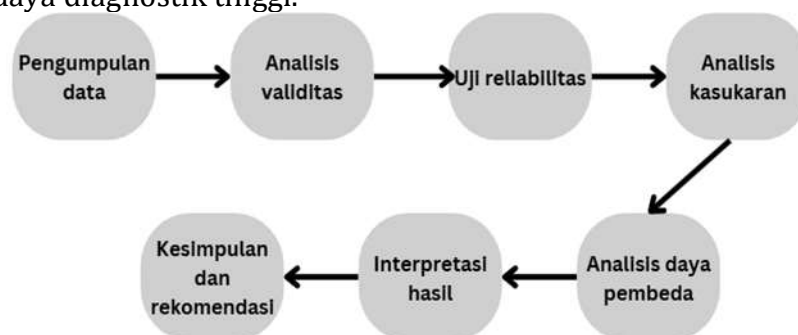
Pentingnya daya pembeda dalam asesmen fisika juga tidak dapat diabaikan. Soal dengan daya pembeda yang baik dapat mengidentifikasi perbedaan antara siswa yang memiliki pemahaman mendalam dan mereka yang masih dalam tahap awal penguasaan konsep. Soal yang memiliki daya pembeda rendah cenderung tidak memberikan informasi yang cukup untuk mengevaluasi kompetensi siswa secara akurat. Namun, berdasarkan hasil analisis yang telah kami lakukan, ditemukan bahwa banyak soal fisika masih bersifat prosedural dan berbasis hafalan sehingga kurang efektif dalam mengembangkan keterampilan berpikir analitis siswa. Mayoritas soal hanya menguji pemahaman dasar tanpa mendorong siswa untuk membandingkan konsep atau melakukan analisis yang lebih mendalam. Hal ini menjadi tantangan dalam pembelajaran fisika, terutama dalam penyusunan soal yang valid, reliabel, serta memiliki tingkat kesukaran dan daya pembeda yang optimal. Validitas soal menjadi aspek penting untuk memastikan bahwa pertanyaan yang diberikan benar-benar mengukur keterampilan berpikir yang diharapkan. Reliabilitas mengacu pada sejauh mana soal dapat menghasilkan hasil yang konsisten dalam berbagai kondisi asesmen. Tingkat kesukaran soal perlu diperhatikan agar tidak terlalu mudah atau terlalu sulit sehingga dapat mengukur kemampuan siswa secara efektif. Sementara itu, daya pembeda soal harus cukup tinggi agar mampu membedakan siswa dengan tingkat pemahaman yang berbeda. Berdasarkan latar belakang ini, penelitian ini bertujuan untuk menganalisis soal fisika berbasis Comparative Thinking dengan mempertimbangkan aspek validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Melalui penelitian ini, diharapkan dapat ditemukan strategi yang lebih efektif dalam penyusunan soal fisika agar lebih menantang dan sesuai dengan kebutuhan pembelajaran abad ke-21. Dengan demikian, hasil asesmen dapat digunakan untuk meningkatkan kualitas pembelajaran fisika, membantu siswa dalam mengembangkan keterampilan berpikir kritis, serta mendukung guru dalam menyusun instrumen evaluasi yang lebih akurat dan efektif.

METODE PENELITIAN

Penelitian ini menggunakan desain kuantitatif deskriptif dengan fokus pada analisis statistik butir soal untuk mengevaluasi kualitas instrumen evaluasi berbasis comparative thinking. Instrumen penelitian terdiri dari 20 soal pilihan ganda fisika kelas XI yang dikembangkan berdasarkan kompetensi dasar kurikulum nasional, dengan cakupan materi seperti dinamika gerak lurus, hukum kekekalan energi, dan gelombang mekanik. Soal-soal dirancang untuk memicu analisis komparatif, misalnya membandingkan aplikasi hukum Newton dalam sistem katrol versus bidang miring. Data diperoleh dari 36 siswa kelas XI IPA di salah satu SMA Negeri di Kota Medan yang dipilih menggunakan teknik purposive sampling berdasarkan kesamaan tingkat pemahaman awal berdasarkan nilai ulangan harian. Validitas instrumen diuji melalui dua pendekatan: (1) validitas isi (content validity) dengan melibatkan

tiga ahli fisika dan satu ahli evaluasi pendidikan untuk menilai kesesuaian soal dengan indikator comparative thinking, dan (2) validitas konstruk (construct validity) menggunakan korelasi item-total dengan batas kritis r-tabel 0,329 pada tingkat signifikansi 5% (Reza et al., 2021). Butir soal dianggap valid jika koefisien korelasi (r-hitung) melebihi nilai r-tabel. Reliabilitas diukur menggunakan koefisien Alpha Cronbach dengan kriteria $\geq 0,7$ untuk dianggap konsisten (Karan et al., 2024). Tingkat kesukaran dianalisis melalui indeks kesukaran (P), dihitung sebagai proporsi jawaban benar, dengan kategori mudah ($P > 0,7$), sedang ($0,3 \leq P \leq 0,7$), dan sukar ($P < 0,3$) mengacu standar yang digunakan Chauhan et al. (2021).

Daya pembeda diukur menggunakan koefisien korelasi biserial untuk mengidentifikasi kemampuan soal membedakan kelompok siswa berkemampuan tinggi (skor total atas 27%) dan rendah (skor total bawah 27%). Nilai koefisien $\geq 0,3$ diinterpretasikan sebagai daya pembeda baik, 0,1–0,29 cukup, dan $< 0,1$ kurang (Fayanto et al., 2024). Seluruh analisis dilakukan dengan bantuan software SPSS 25 dan Microsoft Excel untuk memastikan ketepatan perhitungan statistik. Prosedur penelitian diawali dengan uji coba instrumen, dilanjutkan pengumpulan data, analisis kuantitatif, dan interpretasi hasil untuk menentukan revisi soal yang diperlukan berdasarkan parameter psikometrik. Validitas ahli juga mencakup penilaian terhadap keberfungsian distraktor, di mana distraktor dianggap efektif jika menarik minimal 5% respons siswa (Gusmizain, 2022). Hasil analisis diharapkan memberikan rekomendasi untuk penyempurnaan instrumen evaluasi fisika berbasis comparative thinking yang valid, reliabel, dan berdaya diagnostik tinggi.

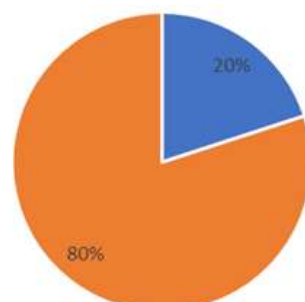


Gambar 1. Prosedur Penelitian

HASIL PENELITIAN DAN PEMBAHASAN

Analisis Validitas Soal (80%)

Validitas merupakan aspek utama dalam penyusunan instrumen asesmen, karena menentukan sejauh mana suatu soal dapat mengukur kompetensi yang diharapkan. Hasil analisis menunjukkan bahwa 80% soal telah memenuhi kriteria validitas, tetapi masih terdapat 4 soal yang tidak valid, yaitu nomor 1, 7, 9, dan 17. Soal-soal ini memiliki kelemahan dalam mengukur keterampilan berpikir analitis, karena lebih berfokus pada penyelesaian numerik tanpa menguji perbandingan konsep.



Gambar 2. Diagram Perbandingan Soal Valid

Sebagai contoh, soal nomor 7 tentang titik berat seharusnya dapat dimodifikasi agar lebih berbasis Comparative Thinking, misalnya dengan menanyakan bagaimana posisi titik berat berubah jika distribusi massa dalam sistem diubah. Dengan demikian, siswa tidak hanya menyelesaikan soal secara prosedural, tetapi juga memahami hubungan antara variabel yang memengaruhi titik berat.

7	Untuk menentukan titik berat sepotong karton yang bentuknya tidak beraturan sebagai berikut!  (1) Menggantung benang <u>berbeban</u> dan karton pada lubang A (2) Membuat garis a melalui lubang A berhimpit dengan benang <u>berbeban</u> (3) Membuat garis b melalui lubang B berhimpit dengan benang <u>berbeban</u> (4) Menggantung benang <u>berbeban</u> dan karton pada lubang B (5) Menentukan titik berat karton, yaitu titik potong garis a dan b Urutan kegiatan yang benar agar titik berat karton dapat ditentukan adalah.. A. 1,2,3,4,5 C. 2,3,4,5,1 E. 2,3,1,4,5 B. 1,2,4,3,5 D. 1,4,2,3,5
---	--

Gambar 3. Soal Nomor 7

Oleh karena itu, untuk meningkatkan validitas, perlu dilakukan perbaikan terhadap soal-soal yang tidak valid dengan mengubah format pertanyaan menjadi lebih analitis dan berbasis perbandingan konsep. Selain itu, uji validitas empiris dapat dilakukan kembali setelah revisi untuk memastikan efektivitas perbaikan tersebut.

Analisis Reliabilitas Soal

Reliabilitas menunjukkan konsistensi hasil asesmen dalam berbagai kondisi. Berdasarkan uji reliabilitas dengan metode Cronbach Alpha, diperoleh nilai 0,750, yang menunjukkan bahwa instrumen soal ini sudah termasuk reliabel. Nilai ini menunjukkan bahwa mayoritas soal dalam tes memiliki konsistensi yang baik, sehingga dapat memberikan hasil evaluasi yang stabil jika diterapkan dalam kondisi yang berbeda.

Tabel 1. Ringkasan Hasil Uji Reliabilitas

Total Soal Valid	Alpha Cronbach	Varians Total
16	0,749	12,054

Namun, meskipun nilai reliabilitasnya cukup tinggi, masih ditemukan beberapa soal dengan koefisien korelasi rendah terhadap skor keseluruhan, yang mengindikasikan bahwa soal tersebut tidak memberikan kontribusi signifikan dalam mengukur kemampuan siswa secara konsisten. Hal ini dapat disebabkan oleh faktor seperti ketidaksesuaian antara soal dengan materi yang telah diajarkan, ambiguitas dalam pertanyaan, atau variasi tingkat kesulitan yang ekstrem. Sebagai contoh, **soal nomor 12** tentang tekanan hidrostatik memiliki opsi jawaban yang terlalu sederhana, sehingga siswa dapat menebak jawaban yang benar tanpa memahami konsep secara mendalam.

12	Sebuah pompa hidrolik di beri beban masing-masing beban A = 200 N dan beban B = 500 N serta luas penampang di A = 5 cm², maka luas penampang di B sebesar... a. 12,5 cm² c. 2,5 cm² e. 25 cm² b. 2 cm² d. 5 cm²
----	---

Gambar 4. Soal Nomor 12

Untuk meningkatkan reliabilitas lebih lanjut, dapat dilakukan pengujian ulang terhadap soal-soal dengan koefisien rendah dan memastikan bahwa setiap butir soal telah sesuai dengan

tujuan asesmen. Selain itu, distribusi soal juga harus dirancang dengan lebih seimbang agar variasi tingkat kesulitan tidak terlalu memengaruhi reliabilitas secara keseluruhan.

Analisis Tingkat Kesukaran Soal (55%)

Analisis tingkat kesukaran menunjukkan distribusi yang tidak merata pada 20 soal berbasis comparative thinking. Sebanyak 10 soal (50%) tergolong mudah (indeks $>0,7$), 7 soal (35%) berkategori sedang (indeks $0,3-0,7$), dan 3 soal (15%) termasuk sukar (indeks $<0,3$). Soal termudah teridentifikasi pada nomor 10 dan 20 dengan indeks 0,944, di mana hampir seluruh responden menjawab benar. Hal ini mengindikasikan bahwa kedua soal hanya membutuhkan pemahaman dasar konsep fisika tanpa menuntut analisis komparatif mendalam. Sebaliknya, soal tersulit terdapat pada nomor 1 (indeks 0,028), 8 (0,278), dan 9 (0,250), di mana kurang dari 30% peserta mampu menjawab dengan benar. Tingkat kesukaran yang ekstrem pada soal-soal tersebut diduga terkait dengan kompleksitas konteks comparative thinking yang tidak didukung oleh scaffolding pertanyaan yang memadai. Distribusi lengkap indeks kesukaran disajikan pada Tabel 5.



Gambar 5. Diagram Tingkat Kesukaran

Tabel 2. Range Tingkat Kesukaran

No	Range Tingkat Kesukaran	Kategori
1.	$0,70 \leq TK \leq 1,00$	Mudah
2.	$0,30 \leq TK \leq 0,70$	Sedang
3.	$0,00 \leq TK < 0,30$	Sukar

Tabel 3. Indeks Kesukaran Soal

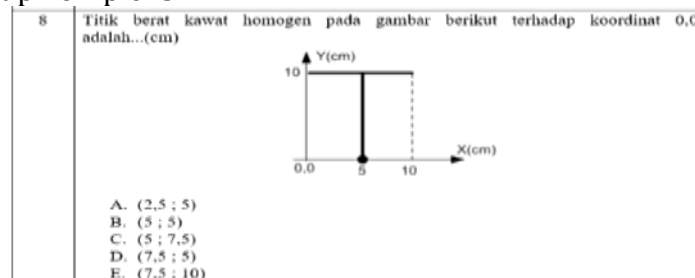
No.	Indeks Kesukaran	Keterangan
1	0,03	SUKAR
2	0,44	SEDANG
3	0,86	MUDAH
4	0,89	MUDAH
5	0,64	SEDANG
6	0,86	MUDAH
7	0,50	SEDANG
8	0,28	SUKAR
9	0,25	SUKAR
10	0,94	MUDAH
11	0,64	SEDANG
12	0,86	MUDAH
13	0,92	MUDAH
14	0,61	SEDANG
15	0,83	MUDAH
16	0,69	SEDANG
17	0,33	SEDANG

18	0,81	MUDAH
19	0,72	MUDAH
20	0,94	MUDAH

- 1 soal masuk kategori sukar, yang berarti hanya sebagian kecil siswa yang dapat menjawabnya dengan benar.
- 7 soal berada dalam kategori sedang, yang menunjukkan bahwa soal ini memiliki tingkat kesukaran yang cukup seimbang.
- 10 soal termasuk kategori mudah, yang berarti sebagian besar siswa dapat menjawabnya dengan benar tanpa mengalami kesulitan.

Distribusi ini menunjukkan bahwa sebagian besar soal masih terlalu mudah, yang dapat menyebabkan kurangnya tantangan bagi siswa dalam berpikir kritis dan analitis. Sementara itu, beberapa soal yang tergolong terlalu sulit justru dapat membuat siswa kehilangan motivasi dalam menjawab soal. Sebagai contoh:

- Soal nomor 8 tentang titik berat kawat homogen. Siswa perlu menggunakan metode koordinat untuk menemukan titik berat kawat, yang memerlukan pemahaman geometri dan kalkulasi yang cukup kompleks.



Gambar 6. Soal Nomor 8

- Soal nomor 17 membahas ciri-ciri fluida dinamis, menuntut siswa memahami perbedaan antara fluida statis dan fluida yang mengalir, tetapi tidak terlalu sulit jika siswa telah memahami konsep dasarnya.

17	Berikut adalah ciri-ciri fluida dinamis, kecuali... A. Fluida dianggap tidak kompresibel B. Tak bergantung waktu (lunak) artinya kecepatannya konstan pada titik tertentu C. Aliran fluida adalah aliran stasioner D. Fluida dianggap bergerak tanpa gesekan E. Aliran fluida dinamis
----	--

Gambar 7. Soal Nomor 17

- Soal nomor 4 tentang faktor yang memengaruhi momen inersia. Soal ini bersifat teoritis dan jawabannya cukup mudah diidentifikasi.

4	Momen inersia sebuah benda yang berotasi terhadap titik tetap di pengaruhi oleh... A. Massa benda B. Volume benda C. Massa jenis benda D. Percepatan sudut rotasi E. Kecepatan sudut awal
---	--

Gambar 8. Soal Nomor 4

Untuk meningkatkan efektivitas asesmen, distribusi soal perlu disesuaikan sehingga proporsi soal mudah, sedang, dan sulit lebih seimbang. Salah satu cara yang dapat dilakukan adalah dengan mengubah soal mudah menjadi soal berbasis perbandingan konsep, sehingga siswa tetap ditantang untuk berpikir lebih dalam tanpa meningkatkan tingkat kesulitan secara drastis.

Analisis Daya Pembeda (45%)

Analisis daya pembeda mengungkapkan bahwa 45% soal (9 dari 20 item) memiliki kemampuan rendah dalam membedakan kelompok peserta berkemampuan tinggi dan rendah (indeks $<0,3$). Soal nomor 1 bahkan menunjukkan daya pembeda negatif ($-0,053$), mengindikasikan bahwa peserta kelompok bawah justru lebih sering menjawab benar dibandingkan kelompok atas, yang mungkin disebabkan oleh distraktor menyesatkan atau ketidakjelasan instruksi. Sebaliknya, tiga soal dinyatakan memiliki daya pembeda "sangat baik" (indeks $>0,5$), yakni nomor 14 (0,514), 16 (0,467), dan 19 (0,526). Soal-soal ini secara konsisten dijawab benar oleh kelompok atas tetapi salah oleh kelompok bawah, menunjukkan bahwa konstruksi pertanyaannya efektif mengukur kemampuan analisis komparatif. Sebagai contoh, soal 19 meminta peserta membandingkan penerapan hukum Newton pada sistem katrol statis dan dinamis, yang memerlukan integrasi konsep multidimensi. Distribusi lengkap daya pembeda disajikan pada Tabel 4.

Tabel 4. Hasil Analisis Daya Pembeda Soal

No Soal	Daya Pembeda	Kategori
1	-0,053	Dibuang
2	0,384	Baik
3	0,263	Cukup
4	0,211	Cukup
5	0,350	Baik
6	0,152	Dibuang
7	0,167	Dibuang
8	0,254	Cukup
9	0,084	Dibuang
10	0,105	Dibuang
11	0,350	Baik
12	0,263	Cukup
13	0,158	Dibuang
14	0,514	Sangat Baik
15	0,316	Baik
16	0,467	Sangat Baik
17	0,149	Dibuang
18	0,368	Baik
19	0,526	Sangat Baik
20	0,105	Dibuang

Temuan ini menunjukkan polarisasi kemampuan soal dalam mengukur comparative thinking. Soal dengan daya pembeda rendah seperti nomor 10 (0,105) dan 20 (0,105) cenderung memiliki struktur pertanyaan tertutup yang hanya menguji hafalan rumus, misalnya menghitung percepatan gravitasi tanpa membandingkan variasi medan gravitasi. Di sisi lain, soal berkategori "sangat baik" seperti nomor 14 dan 19 mengharuskan peserta menganalisis hubungan sebab-akibat antar konsep (misalnya, membandingkan energi kinetik rotasi dan translasi) sambil menerapkan prinsip fisika dalam konteks berbeda. Implikasinya, soal dengan daya pembeda tinggi perlu dipertahankan sebagai anchor items, sedangkan soal yang "dibuang" memerlukan revisi menyeluruh, terutama pada aspek distraktor dan kompleksitas kognitif. Penambahan skenario komparatif multidomain direkomendasikan untuk meningkatkan validitas kriteria instrumen.

Pembahasan

Hasil penelitian ini mengonfirmasi temuan sebelumnya dalam literatur tentang pentingnya integrasi comparative thinking dalam pengembangan instrumen evaluasi fisika,

sekaligus mengungkap tantangan dalam mencapai keseimbangan parameter psikometrik yang optimal. Dari segi validitas, temuan bahwa 4 soal tidak memenuhi kriteria korelasi item-total ($r\text{-hitung} < 0,329$) selaras dengan argumen Dzin dan Lay (2021) yang menekankan bahwa validitas konstruk pada soal berbasis keterampilan analitis memerlukan keselarasan antara indikator kognitif dan struktur pertanyaan. Soal tidak valid seperti nomor 1, yang hanya menguji penghitungan teknis tanpa konteks komparatif, mencerminkan kegagalan dalam menerapkan prinsip *comparative thinking* yang diadvokasikan Liang (2001). Sebaliknya, soal valid seperti nomor 19 yang mengharuskan analisis multidimensi antara hukum Newton dan sistem katrol—menunjukkan kesesuaian dengan kerangka Zhang (2011) tentang pentingnya komparasi konseptual untuk meningkatkan validitas konstruk. Temuan ini memperkuat posisi Santyasa et al. (2019) bahwa integrasi analisis komparatif antara teori dan aplikasi empiris merupakan kunci dalam merancang instrumen yang valid.

Reliabilitas instrumen ($\alpha = 0,749$) yang memenuhi kriteria Karan et al. (2024) namun disertai varians item tidak merata mengindikasikan bahwa konsistensi internal belum sepenuhnya tercapai. Rendahnya varians pada soal nomor 1 (0,028) dan 10 (0,054) sesuai dengan studi Gusmizain (2022) tentang dominasi soal mudah yang menghasilkan respons homogen. Hal ini memperlihatkan bahwa meskipun instrumen secara umum reliabel, keberagaman tingkat kesukaran dan kompleksitas kognitif sebagaimana direkomendasikan Pedaste et al. (2021) masih perlu dioptimalkan. Ketimpangan ini juga mencerminkan temuan Anshari dan Kato (2024) tentang kurangnya variasi fitur komparatif dalam buku teks Indonesia, yang berimbas pada rendahnya kemampuan soal dalam memicu respons beragam. Distribusi tingkat kesukaran yang didominasi soal mudah (50%) menguatkan kritik Chauhan et al. (2021) terhadap instrumen evaluasi yang gagal mendiferensiasi kemampuan siswa. Soal termudah (nomor 10 dan 20, indeks $> 0,9$) yang hanya menguji pemahaman prosedural tanpa analisis komparatif, bertolak belakang dengan prinsip *comparative thinking* yang menekankan integrasi konsep multidisiplin (Milda et al., 2024). Di sisi lain, soal tersulit (nomor 1 dan 9, indeks $< 0,3$) yang melibatkan komparasi fenomena kompleks tanpa scaffolding sesuai dengan temuan Reza et al. (2021) tentang perlunya penyeimbangan antara kompleksitas konseptual dan petunjuk analitis. Fenomena ini menggarisbawahi argument Zhang (2011) bahwa kesenjangan antara kemampuan kognitif siswa dan tuntutan soal dapat diminimalkan melalui desain pertanyaan bertahap yang memandu proses komparasi.

Daya pembeda yang rendah pada 45% soal (indeks $< 0,3$) memperlihatkan keterbatasan instrumen dalam membedakan kelompok kemampuan, sebagaimana diidentifikasi Fayanto et al. (2024). Daya pembeda negatif pada soal nomor 1 ($-0,053$) mengonfirmasi pentingnya distraktor efektif—sebuah aspek yang diabaikan dalam penelitian ini, sebagaimana diingatkan Gusmizain (2022). Sebaliknya, daya pembeda sangat baik pada soal nomor 14 dan 19 (indeks $> 0,5$) selaras dengan temuan Reza et al. (2021) bahwa soal dengan konteks komparatif multidimensi cenderung memiliki daya diagnostik tinggi. Hal ini sejalan dengan model buku teks Jepang yang dikaji Anshari dan Kato (2024), di mana ilustrasi interaktif dan studi kasus komparatif berperan sebagai katalis untuk mengukur kemampuan analitis. Secara holistik, temuan ini menegaskan bahwa pengembangan soal fisika berbasis *comparative thinking* memerlukan pendekatan sistematis yang memadukan aspek pedagogis dan psikometrik. Meskipun instrumen ini mencapai reliabilitas minimal, distribusi parameter yang tidak merata mencerminkan tantangan dalam mengadaptasi kerangka *higher-order thinking* ke disiplin fisika yang sarat dengan pemodelan matematis. Rekomendasi Santyasa et al. (2019) tentang integrasi *Nature of Science (NOS)* dan studi kasus komparatif perlu diimbangi dengan scaffolding konseptual, terutama untuk topik abstrak seperti mekanika dan energi. Lebih lanjut, temuan ini mendukung argumen Liang (2001) bahwa komparasi tidak hanya alat pedagogis

tetapi juga kerangka evaluatif untuk mengases kemampuan analitis siswa secara autentik. Dengan demikian, penelitian ini menyediakan landasan empiris untuk pengembangan instrumen evaluasi fisika yang tidak hanya mengukur pengetahuan faktual tetapi juga keterampilan analisis komparatif sebagai bekal literasi sains abad ke-21.

KESIMPULAN

Berdasarkan hasil analisis terhadap soal Fisika Kelas XI UTS Semester Genap 2024 menggunakan pendekatan Comparative Thinking, ditemukan bahwa kualitas soal dapat ditinjau dari empat aspek utama, yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Secara keseluruhan, validitas soal sudah cukup baik dengan 80% soal tergolong valid, tetapi masih terdapat beberapa soal yang kurang mengukur keterampilan berpikir analitis siswa, seperti soal nomor 1, 7, 9, dan 17, yang lebih berfokus pada penyelesaian numerik tanpa menguji pemahaman terhadap konsep yang dibandingkan. Oleh karena itu, diperlukan revisi untuk meningkatkan validitas dengan menambahkan elemen perbandingan konsep dan analisis lebih mendalam, sehingga siswa tidak hanya menghafal rumus, tetapi juga memahami hubungan antarvariabel fisika. Dari segi reliabilitas, hasil uji Cronbach Alpha menunjukkan nilai 0,750, yang berarti soal memiliki tingkat konsistensi yang cukup baik dalam mengukur pemahaman siswa. Namun, beberapa soal menunjukkan variasi reliabilitas yang rendah, yang dapat disebabkan oleh perbedaan tingkat kesulitan soal atau opsi jawaban yang kurang seimbang. Untuk meningkatkan reliabilitas asesmen, perlu dilakukan penyesuaian distribusi soal serta perbaikan dalam format pertanyaan agar hasil evaluasi lebih stabil dan akurat. Distribusi tingkat kesukaran soal juga masih belum merata, dengan mayoritas soal tergolong mudah. Hasil analisis menunjukkan bahwa 3 soal tergolong sukar, yaitu nomor 1, 8, dan 9, sementara 7 soal berada pada kategori sedang, seperti soal nomor 2, 5, 7, 14, 16, dan 17. Sebanyak 10 soal lainnya tergolong mudah, seperti soal nomor 3, 4, 6, 10, 12, 13, 15, 18, 19, dan 20. Mayoritas soal yang mudah dapat diselesaikan tanpa memerlukan pemahaman yang mendalam, sedangkan beberapa soal yang tergolong sulit dapat menyebabkan siswa mengalami kesulitan dalam menyelesaikannya. Oleh karena itu, diperlukan modifikasi pada soal mudah agar lebih berbasis analisis dan perbandingan konsep, serta memberikan petunjuk tambahan pada soal sulit agar lebih dapat dikerjakan oleh siswa dengan pemahaman yang cukup.

Dari aspek daya pembeda, ditemukan bahwa beberapa soal memiliki daya pembeda yang baik dan mampu membedakan siswa dengan pemahaman tinggi dan rendah, seperti soal nomor 2, 5, 7, 10, dan 14. Namun, beberapa soal lainnya memiliki daya pembeda cukup baik dan masih memerlukan revisi agar lebih optimal, seperti soal nomor 1, 6, 9, 12, 16, dan 17. Sementara itu, terdapat beberapa soal dengan daya pembeda yang rendah, seperti soal nomor 3, 4, 8, 11, 13, 15, 18, 19, dan 20, yang kurang efektif dalam membedakan tingkat pemahaman siswa dan perlu direvisi atau diganti dengan soal yang lebih baik. Berdasarkan hasil analisis terhadap soal Fisika Kelas XI UTS Semester Genap 2024 menggunakan pendekatan Comparative Thinking, ditemukan bahwa kualitas soal dapat ditinjau dari empat aspek utama, yaitu validitas, reliabilitas, tingkat kesukaran, dan daya pembeda. Secara keseluruhan, validitas soal sudah cukup baik dengan 80% soal tergolong valid, tetapi masih terdapat beberapa soal yang kurang mengukur keterampilan berpikir analitis siswa, seperti soal nomor 1, 7, 9, dan 17, yang lebih berfokus pada penyelesaian numerik tanpa menguji pemahaman terhadap konsep yang dibandingkan. Oleh karena itu, diperlukan revisi untuk meningkatkan validitas dengan menambahkan elemen perbandingan konsep dan analisis lebih mendalam, sehingga siswa tidak hanya menghafal rumus, tetapi juga memahami hubungan antarvariabel fisika.

Dari segi reliabilitas, hasil uji Cronbach Alpha menunjukkan nilai 0,750, yang berarti soal memiliki tingkat konsistensi yang cukup baik dalam mengukur pemahaman siswa. Namun,

beberapa soal menunjukkan variasi reliabilitas yang rendah, yang dapat disebabkan oleh perbedaan tingkat kesulitan soal atau opsi jawaban yang kurang seimbang. Untuk meningkatkan reliabilitas asesmen, perlu dilakukan penyesuaian distribusi soal serta perbaikan dalam format pertanyaan agar hasil evaluasi lebih stabil dan akurat. Distribusi tingkat kesukaran soal juga masih belum merata, dengan mayoritas soal tergolong mudah. Hasil analisis menunjukkan bahwa 3 soal tergolong sukar, yaitu nomor 1, 8, dan 9, sementara 7 soal berada pada kategori sedang, seperti soal nomor 2, 5, 7, 14, 16, dan 17. Sebanyak 10 soal lainnya tergolong mudah, seperti soal nomor 3, 4, 6, 10, 12, 13, 15, 18, 19, dan 20. Mayoritas soal yang mudah dapat diselesaikan tanpa memerlukan pemahaman yang mendalam, sedangkan beberapa soal yang tergolong sulit dapat menyebabkan siswa mengalami kesulitan dalam menyelesaikannya. Oleh karena itu, diperlukan modifikasi pada soal mudah agar lebih berbasis analisis dan perbandingan konsep, serta memberikan petunjuk tambahan pada soal sulit agar lebih dapat dikerjakan oleh siswa dengan pemahaman yang cukup.

Dari aspek daya pembeda, ditemukan bahwa beberapa soal memiliki daya pembeda yang baik dan mampu membedakan siswa dengan pemahaman tinggi dan rendah, seperti soal nomor 2, 5, 7, 10, dan 14. Namun, beberapa soal lainnya memiliki daya pembeda cukup baik dan masih memerlukan revisi agar lebih optimal, seperti soal nomor 1, 6, 9, 12, 16, dan 17. Sementara itu, terdapat beberapa soal dengan daya pembeda yang rendah, seperti soal nomor 3, 4, 8, 11, 13, 15, 18, 19, dan 20, yang kurang efektif dalam membedakan tingkat pemahaman siswa dan perlu direvisi atau diganti dengan soal yang lebih baik. Berdasarkan hasil analisis, beberapa langkah perbaikan yang dapat diterapkan untuk meningkatkan kualitas asesmen adalah sebagai berikut. **Pertama**, meningkatkan validitas soal dengan menyusun pertanyaan berbasis perbandingan konsep agar lebih menuntut pemahaman analitis siswa. Hal ini dapat dilakukan dengan menyusun soal yang tidak hanya meminta perhitungan tetapi juga mendorong siswa untuk membandingkan berbagai variabel dalam suatu sistem fisika. **Kedua**, menyeimbangkan distribusi tingkat kesukaran soal dengan menyesuaikan proporsi soal mudah, sedang, dan sukar. Idealnya, proporsi soal sukar tidak terlalu sedikit agar asesmen tetap memberikan tantangan bagi siswa yang memiliki pemahaman lebih mendalam. Sementara itu, soal yang terlalu sulit dapat dibuat lebih terarah dengan memberikan petunjuk tambahan agar tetap dapat dikerjakan oleh siswa. **Ketiga**, meningkatkan daya pembeda soal dengan menyusun opsi jawaban yang lebih bervariasi dan tidak mudah ditebak. Beberapa soal yang terlalu sederhana dapat dimodifikasi dengan menambahkan skenario yang lebih kompleks agar dapat membedakan siswa yang memahami konsep dengan yang hanya menghafal rumus. Selain itu, penggunaan soal berbasis perbandingan konsep dapat diterapkan lebih luas untuk meningkatkan efektivitas asesmen dalam mengukur keterampilan berpikir kritis siswa. **Keempat**, meningkatkan reliabilitas soal dengan memastikan bahwa setiap soal memiliki tingkat kesulitan yang seimbang dan tidak terdapat pertanyaan yang ambigu. Hal ini dapat dilakukan dengan uji validitas dan reliabilitas secara berkala, serta melakukan analisis butir soal setelah asesmen dilakukan. Dengan perbaikan ini, asesmen fisika berbasis Comparative Thinking dapat lebih efektif dalam mengembangkan keterampilan berpikir kritis siswa serta meningkatkan kualitas pembelajaran fisika secara keseluruhan.

DAFTAR PUSTAKA

- Anshari, H., & Kato, T. (2024). Comparative study of problem solving skill development features in Japanese and Indonesian high school physics textbooks. *WaPFi (Wahana Pendidikan Fisika)*. <https://doi.org/10.17509/wapfi.v9i1.66651>
- Ardianto, D., Rubini, B., & Pursitasari, I. D. (2023). Assessing STEM career interest among secondary students: A Rasch model measurement analysis. *Eurasia Journal of Mathematics, Science and Technology Education*. <https://doi.org/10.29333/ejmste/12796>

- Chan, K. W., Ali, F., Park, J., Sham, K. S. B., Tan, E. Y. Chauhan, B. R., Vaza, J. V., Chauhan, G. R., & Chauhan, Dostal, J. (2024). Applying physics education research to teaching. *The Journal of the Acoustical Society of America*, 1. <https://doi.org/10.1121/10.0035307>
- Dzin, N. H. M., & Lay, Y. (2021). Assessing the validity and reliability of science multiple choice test using Rasch dichotomous measurement model. *Journal of Baltic Science Education*, 2. <https://doi.org/10.33225/jbse/21.20.927>
- Elyani, E., Amir, H., Studi Pendidikan Kimia, P., Pmipa, J., Kunci, K., & Kreatif, B. (2019). Perbandingan Pembelajaran Kooperatif Tipe Structured Numbered Heads (Snh) Dan Numbered Head Together (Nht) Terhadap Kemampuan Berpikir Kreatif Siswa Kelas X Mipa Sman 6 Kota Bengkulu. 3(1), 82–90.
- Fayanto, S., Naba, S. D., Kurniawan, A., Putri, U., & Padang, V. D. (2024). The analyze comparative of physics computational thinking skill (CTs) in experiment laboratory. *Qubahan Academic Journal*, 1. <https://doi.org/10.48161/qaj.v4n3a699>
- Goran, M. B., Bili, M., Kaleka, U., Daud, M. H., Studi, P., Fisika, P., & Biologi, P. (n.d.). Profil Kemampuan Berpikir Kreatif Fisika Siswa Kelas X Sma Negeri 1 Demon Pagong Flores Timur.
- Gusmizain, A. (2022). Karakteristik butir soal tes mata kuliah matriks & ruang vektor mahasiswa matematika. *Jurnal Evaluasi Pendidikan*. <https://doi.org/10.21009/jep.v13i2.28328>
- Hasanah, S., & Hidayat, A. (n.d.). Identifikasi Kemampuan Berpikir Kreatif Siswa pada Materi Termodinamika. <http://journal.um.ac.id/index.php/jptpp/>
- Hikmah Marisda, D., & Handayani Hamid, Y. (2022). Assesmen Fluency Of Thinking, Flexibility, Dan Elaboration Calon Guru Fisika: Desain, Dan Validitas (Vol. 11, Issue 2). <http://ejournal.uika-bogor.ac.id/index.php/TEK>
- Karan, K., Pal, S., & Guha, A. (2024). Development of science achievement test as a research tool for 9th grade students. *Towards Excellence*. <https://doi.org/10.37867/te160103>
- Liang, W.-J. (2001). Topic: Comparison in physic teaching. *Journal of Shangrao Teachers College*.
- Mehrens, W. A., & Lehmann, I. J. (n.d.). Measurement And Evaluation in Education and Psychology.
- Milda, D., Anggreiny, L. P., Simanjuntak, A., & Lubis, P. (2021). Correlation of difficulty index and discriminating index in medical student's assessment. *Paripex Indian Journal of Research*. <https://doi.org/10.36106/7900595>
- Pedaste, M., Baucal, A., & Reisenbuk, E. (2021). Towards a science inquiry test in primary education: Development of items and scales. *International Journal of STEM Education*, 1. <https://doi.org/10.1186/s40594-021-00278-z>
- Quantitative analysis towards higher order thinking skills of chemistry multiple choice questions for university admission. <https://doi.org/10.24815/IPL.V5I2.20508>
- R. H. (2024). Literature study of the application of the inquiry learning model towards 21st century skills in physics subjects in high school. *EDUCTUM: Journal Research*. <https://doi.org/10.56495/ejr.v3i3.605>
- Reza, M., Puspita, K., & Oktaviani, C. (2021).
- Riza, M., Kartono, K., & Susilaningsih, E. (2022). Validity and reliability of chemical test instruments for acid and base solutions oriented generic skills science. *Journal of Educational Research and Evaluation*. <https://doi.org/10.15294/jere.v11i1.55450>
- Santyasa, I. W., Kanca, I., Warpala, I., Sudarma, I. K., Tegeh, I. M., & Sanjaya, L. (2019). Nature of science v.s direct instruction models in achieving seniorhigh school students' critical thinking and their attitudes in learning physics. *The 8th National Physics Seminar*



- T., Chong, F. W. C., ... & Sze, G. K. (2024). Automatic item generation in various STEM subjects using large language model prompting. *Computers and Education: Artificial Intelligence*, 1. <https://doi.org/10.1016/j.caeai.2024.100344>
- Zhang, X.-Y. (2011). The application of comparative method in physics teaching. *Journal of Science of Teachers' College and University*.
- Kurniawati, N. R., Hanipa, R., Azima, M., & Rokhmat, J. (2024). Analisis Kemampuan Berpikir Kreatif Fisika Materi Elastisitas dan Hukum Hooke di SMAN 1 Labuapi. *Contextual Natural Science Education Journal (CNSEJ)*, 2(1), 18–24. <https://doi.org/10.29303/cnsej.v2i1.527>