Inovasi Pendidikan Fisika
p-ISSN: 2302-4496
e-ISSN: 2830-3881

Vol 15, No 01, 121-131
https://doi.org/10.26740/ipf
jipf@unesa.ac.id

Pengembangan dan Analisis Kualitas Instrumen Asesmen Digital Keterampilan Berpikir Kritis pada Materi Energi Terbarukan

Graciella Setiarini[1] dan Abu Zainuddin[2,*]
[1,2]Program Studi Pendidikan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Surabaya, Surabaya, Indonesia
Email: graciella.22096@mhs.unesa.ac.id abuzainudin@unesa.ac.id

Abstrak

Perkembangan teknologi pendidikan mendorong pemanfaatan asesmen digital untuk mengukur keterampilan berpikir kritis, tetapi *Quizizz* dan *Google Form* di sekolah masih sebatas media pelaksanaan tes dan belum didukung instrumen yang teruji kualitasnya. Penelitian ini bertujuan mengembangkan instrumen asesmen digital keterampilan berpikir kritis pada materi energi terbarukan menggunakan *Quizizz* dan *Google Form*, serta menganalisis kualitas instrumen dan kepraktisan penggunaannya. Penelitian pengembangan (*Research and Development*) ini memakai model ADDIE (*Analyze, Design, Development, Implement,* dan *Evaluation*). Validasi melibatkan tiga validator (dua dosen ahli dan satu guru fisika), uji coba terbatas melibatkan 33 peserta didik, dan implementasi melibatkan 64 peserta didik kelas X-2 dan X-3 SMA Katolik Karitas III Surabaya. Data dikumpulkan melalui lembar validasi, tes, dan angket respons, kemudian dianalisis menggunakan persentase validitas, *Percentage of Agreement* (PoA), korelasi *Product Moment*, KR-20, tingkat kesukaran, daya pembeda, dan persentase respons. Instrumen yang dihasilkan terdiri atas 20 butir pilihan ganda berkategori sangat valid (93,75%) dengan kesepakatan antar validator yang tinggi (PoA 95,24%). Seluruh butir valid secara empiris (r hitung 0,433–0,700 > r tabel 0,344), reliabilitas sebesar 0,87 (sangat tinggi), 18 butir berkategori sedang dan 2 butir sukar, serta daya pembeda 0,44–0,89 (baik sampai sangat baik). Kepraktisan berkategori sangat praktis, yaitu 94,98% pada *Quizizz* dan 94,95% pada *Google Form*. Instrumen ini dapat digunakan guru sebagai alternatif asesmen berpikir kritis yang siap pakai pada dua *platform*, dan prosedur pengembangannya dapat menjadi acuan bagi pengembangan instrumen serupa pada materi fisika lain.

Kata kunci: asesmen digital, keterampilan berpikir kritis, energi terbarukan, *Quizizz*, *Google Form*, ADDIE

Development and Quality Analysis of a Digital Assessment Instrument for Critical Thinking Skills on Renewable Energy

Abstract

The development of educational technology has encouraged the use of digital assessment to measure critical thinking skills. However, platforms such as Quizizz *and* Google Form *in schools are still used mainly to deliver tests and are not supported by instruments of proven quality. This study aimed to develop a digital assessment instrument for critical thinking skills on renewable energy using* Quizizz *and* Google Form*, and to analyze the instrument quality and the practicality of the platforms. This research and development study used the ADDIE model (*Analyze, Design, Development, Implement, *and* Evaluation*). Three validators (two lecturers and one physics teacher) validated the instrument, a limited trial involved 33 students, and the implementation involved 64 tenth-grade students of SMA Katolik Karitas III Surabaya. Data were collected using validation sheets, a test, and response questionnaires, and analyzed using validity percentage,*

Copyright © The Author(s). This work is licensed under CC BY-NC 4.0

Percentage of Agreement (PoA), Pearson product-moment correlation, KR-20, item difficulty, item discrimination, and response percentage. The instrument consisted of 20 multiple-choice items and was categorized as very valid (93.75%) with high agreement among validators (PoA 95.24%). All items were empirically valid (r-count 0.433–0.700 > r-table 0.344), reliability was 0.87 (very high), 18 items were of moderate difficulty and 2 items were difficult, and discrimination indices ranged from 0.44 to 0.89 (good to very good). Practicality was categorized as very practical, at 94.98% for Quizizz *and 94.95% for* Google Form*. The instrument can be used by teachers as a ready-to-use alternative for assessing critical thinking on two platforms, and the development procedure can serve as a reference for similar instruments on other physics topics.*

Keywords: *digital assessment, critical thinking skills, renewable energy,* Quizizz*,* Google Form*, ADDIE*

Histori Naskah

Diserahkan: 3 September 2026 Direvisi: 12 Oktober 2026 Diterima: 20 Oktober 2026

How to cite:

Setiarini, G., and Zainuddin, A. (2025). Pengembangan dan Analisis Kualitas Instrumen Asesmen Digital Keterampilan Berpikir Kritis pada Materi Energi Terbarukan. *Inovasi Pendidikan Fisika*, **15**(1), 121-131. DOI: https://doi.org/10.26740/ipf.vxx.nxx.xx-xx.

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi pada era Revolusi Industri 4.0 dan *Society* 5.0 mengubah cara pembelajaran dan evaluasi dilaksanakan. Teknologi tidak lagi berperan sekadar sebagai alat bantu mengajar, tetapi menjadi bagian dari cara guru merancang pengalaman belajar sekaligus menilai hasilnya. Salah satu perubahan yang paling terasa terjadi pada asesmen, yaitu peralihan dari tes berbasis kertas (*paper-based test*) menuju asesmen digital. Safitri dkk. (2023) mengemukakan bahwa penilaian berbasis kertas memiliki sejumlah kelemahan, antara lain kesalahan saat pengoreksian, biaya penggandaan soal, dan waktu pemeriksaan yang kurang efisien, sehingga guru ditantang untuk menyediakan media evaluasi yang lebih baru dan menarik. Asesmen digital menjawab sebagian tantangan tersebut karena skor dapat diolah secara otomatis dan hasil dapat diperoleh dengan cepat, sehingga guru memiliki lebih banyak waktu untuk menafsirkan hasil dan menindaklanjutinya. Balqis dan Andriani (2024), misalnya, melaporkan bahwa fitur penilaian otomatis pada *Quizizz* mempermudah proses evaluasi dan pengolahan hasil belajar.

Namun, kemudahan teknis saja tidak cukup. Perubahan paradigma pendidikan di Indonesia tampak pada Profil Pelajar Pancasila dalam Kurikulum Merdeka, yang menuntut peserta didik merekonstruksi pengalaman belajar berdasarkan kehidupan nyata dan menguatkan kompetensi berpikir kritis (Karira dkk., 2023). Berpikir kritis adalah proses yang logis dan reflektif untuk menentukan apa yang sebaiknya diyakini dan dilakukan (Ennis, 2011). Facione (2015) menjabarkannya ke dalam kemampuan interpretasi, analisis, evaluasi, inferensi, eksplanasi, dan regulasi diri. Dalam kehidupan sehari-hari, kemampuan ini dibutuhkan ketika seseorang harus memilah informasi yang beredar di media sosial dan iklan, membedakan bukti dari opini, mengenali asumsi yang tersembunyi dalam sebuah pernyataan, lalu mengambil keputusan yang dapat dipertanggungjawabkan. Isu energi merupakan contoh yang dekat dengan peserta didik. Klaim bahwa biomassa pasti ramah lingkungan karena berasal dari bahan alami, atau bahwa listrik tenaga surya terlalu mahal dan tidak stabil, hanya dapat dinilai dengan benar oleh orang yang mampu membaca data, memeriksa kredibilitas sumber, dan menimbang bukti secara seimbang. Dalam pelajaran fisika, kemampuan yang sama dilatihkan melalui analisis grafik dan tabel, penarikan simpulan dari hasil pengamatan, serta evaluasi terhadap klaim ilmiah.

Sayangnya, keterampilan berpikir kritis peserta didik di Indonesia masih tergolong rendah. Karira dkk. (2023) merangkum sejumlah penelitian yang menunjukkan ketercapaian berpikir kritis peserta didik berada pada kategori rendah hingga sangat rendah, dan menjelaskan bahwa salah satu penyebabnya adalah peserta didik tidak terbiasa mengerjakan soal yang memuat keterampilan berpikir kritis. Guru fisika juga disebut belum pernah melakukan penilaian khusus terhadap keterampilan tersebut, dan asesmen yang disusun guru cenderung berfokus pada penguasaan konsep tingkat hafalan. Kondisi serupa ditemukan melalui wawancara dengan guru fisika di SMA Katolik Karitas III Surabaya. Guru telah memanfaatkan *Quizizz* dan *Google Form*, tetapi penggunaannya masih diarahkan pada kebutuhan administratif atau sebagai alat evaluasi yang interaktif, bukan untuk mengukur keterampilan berpikir kritis secara khusus. Guru juga menyampaikan bahwa kedua *platform* tersebut rentan terhadap kecurangan, misalnya peserta didik membuka tab baru untuk mencari jawaban. Akibatnya, hasil asesmen belum menggambarkan kemampuan analisis dan evaluasi peserta didik yang sesungguhnya.

Keterbatasan tersebut menunjukkan bahwa yang dibutuhkan bukan hanya media yang menarik, tetapi instrumen yang kualitasnya terbukti. Instrumen yang baik harus memenuhi kelayakan teoretis (menurut ahli) dan kelayakan empiris (hasil uji coba), yang mencakup validitas butir, reliabilitas, tingkat kesukaran, dan daya pembeda (Alifah & Sunarti, 2024; Arikunto, 2018). Kelayakan empiris tidak dapat diabaikan karena butir yang dinilai sangat valid oleh ahli belum tentu valid ketika diujikan kepada peserta didik. Karira dkk. (2023), misalnya, memperoleh validitas teoretis rata-rata 98% pada instrumen berpikir kritis materi energi terbarukan, tetapi hanya sembilan dari 15 butir yang valid secara empiris. Dengan demikian, pengembangan instrumen asesmen digital perlu ditempuh melalui prosedur yang sistematis, misalnya model ADDIE (Branch, 2009), dan diuji secara teoretis maupun empiris sebelum digunakan.

Sejumlah penelitian terdahulu telah menunjukkan potensi asesmen digital berbasis *Quizizz* maupun *Google Form*. Safitri dkk. (2023) mengembangkan instrumen asesmen keterampilan berpikir kritis berbasis *Quizizz* pada pelajaran IPA SMP dengan model ADDIE dan memperoleh instrumen yang valid, praktis, dan efektif. Alifah dan Sunarti (2024) mengembangkan instrumen tes keterampilan berpikir kritis berbasis *Google Form* pada materi Asas Black dan menyimpulkan instrumen tersebut layak digunakan. Azizah dan Hidayat (2024) melaporkan bahwa asesmen digital berbasis *Quizizz* yang mereka kembangkan berkategori sangat valid dan praktis. Meskipun demikian, penelitian-penelitian tersebut menggunakan satu *platform* untuk satu instrumen dan tidak berfokus pada materi energi terbarukan. Penelitian pada materi energi terbarukan oleh Karira dkk. (2023) menggunakan soal uraian berbasis literasi sains, dibatasi sampai tahap pengembangan, dan tidak membahas penyajian pada *platform* digital.

Materi energi terbarukan layak dipilih sebagai konteks pengukuran. Materi ini dipelajari pada kelas X semester satu dalam Kurikulum Merdeka, sangat dekat dengan kehidupan peserta didik, dan berkaitan dengan isu kebutuhan energi yang menjadi pembicaraan global (Karira dkk., 2023). Karakteristiknya yang kontekstual menyediakan banyak stimulus berupa data, grafik, dan kasus nyata, misalnya efisiensi panel surya, perbandingan turbin angin, atau dampak penggunaan biomassa, yang menuntut peserta didik menganalisis dan mengevaluasi, bukan sekadar mengingat.

Mengukur keterampilan berpikir kritis juga menuntut bentuk soal yang tepat. Soal yang hanya meminta peserta didik mengingat definisi atau rumus tidak akan menunjukkan apakah peserta didik mampu menganalisis data atau mengevaluasi suatu klaim. Karena itu, butir soal perlu dirancang pada tingkat kognitif analisis (C4) dan evaluasi (C5), disertai stimulus berupa teks, grafik, tabel, atau gambar yang memuat informasi yang harus ditelaah sebelum peserta didik menjawab. Karira dkk. (2023) menunjukkan bahwa stimulus semacam ini memungkinkan peserta didik menjawab soal berdasarkan data dan bukti yang tersedia. Pada penelitian ini, setiap butir dirancang mengacu pada indikator berpikir kritis Ennis (2011), misalnya menyimpulkan dari data, mempertimbangkan kredibilitas sumber, mengidentifikasi asumsi, dan menganalisis hasil observasi. Rancangan tersebut diharapkan membuat instrumen benar-benar mengukur proses bernalar, bukan sekadar daya ingat.

Pemilihan *Quizizz* dan *Google Form* didasarkan pada karakteristik keduanya yang saling melengkapi. *Quizizz* memanfaatkan unsur permainan, seperti poin dan papan peringkat, yang menurut temuan yang dirangkum Safitri dkk. (2023) membuat pengalaman evaluasi lebih menarik, fleksibel, dan mengurangi rasa bosan. *Google Form* menawarkan penyajian soal yang sederhana, formulir identitas, dan rekapitulasi jawaban otomatis yang memudahkan guru mengolah data, dan telah dipakai sebagai media instrumen tes berpikir kritis dalam fisika (Alifah & Sunarti, 2024). Keduanya sudah akrab bagi guru dan peserta didik, tetapi pemilihan salah satunya sering didasarkan pada kebiasaan, bukan pada bukti tentang kualitas instrumen dan kepraktisan penggunaannya. Jika instrumen yang sama dapat disajikan pada kedua *platform* dengan kualitas yang terjaga, guru memiliki fleksibilitas untuk menyesuaikan media dengan kondisi kelas, ketersediaan perangkat, dan kestabilan koneksi internet.

Pengembangan instrumen semacam ini juga relevan dengan kebutuhan guru. Guru yang telah memahami pentingnya berpikir kritis sering kali belum memiliki contoh instrumen yang siap pakai dan telah diuji kualitasnya, sementara menyusun dan menganalisis butir soal sendiri membutuhkan waktu dan pemahaman teknis. Instrumen yang telah melalui validasi ahli, uji empiris, dan kajian kepraktisan dapat langsung dipakai atau diadaptasi untuk kelas lain, sehingga guru dapat memetakan keterampilan berpikir kritis peserta didik dan menindaklanjutinya dalam pembelajaran. Dalam jangka panjang, pembiasaan peserta didik mengerjakan soal yang menuntut penalaran diharapkan membantu mereka mengambil keputusan yang rasional terhadap isu-isu energi dan lingkungan yang akan mereka hadapi sebagai warga negara.

Kualitas instrumen dan kepraktisan media merupakan dua hal yang berbeda dan sama-sama perlu dikaji. Kualitas instrumen berkaitan dengan ketepatan dan konsistensi pengukuran, sedangkan kepraktisan berkaitan dengan kemudahan penggunaan di kelas. Safitri dkk. (2023) menjelaskan bahwa kepraktisan dipengaruhi oleh lama waktu penggunaan, kemudahan mengadministrasi, dan kemudahan menginterpretasi aplikasi. Instrumen yang sangat valid tetapi sulit dioperasikan akan jarang dipakai guru, sebaliknya media yang mudah digunakan tetapi menyajikan soal yang tidak teruji tidak akan memberi gambaran yang tepat tentang kemampuan peserta didik. Oleh karena itu, penelitian ini mengkaji keduanya sekaligus agar produk yang dihasilkan tidak hanya layak secara ilmiah, tetapi juga dapat diterapkan dalam praktik pembelajaran sehari-hari.

Penelitian ini juga menambahkan satu aspek yang belum selalu dilaporkan pada penelitian pengembangan instrumen, yaitu kesepakatan antar validator. Persentase validitas menunjukkan seberapa tinggi skor yang diberikan ahli, tetapi belum menunjukkan apakah para ahli memberikan penilaian yang sejalan. Ukuran kesepakatan seperti *Percentage of Agreement* (Borich, 1994) melengkapi persentase validitas karena menunjukkan bahwa kesimpulan tentang kelayakan instrumen tidak bergantung pada pendapat satu orang saja. Dengan demikian, kualitas instrumen ditelaah dari sudut pandang ahli (validitas teoretis dan kesepakatan antar validator) sekaligus dari sudut pandang data peserta didik (validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda).

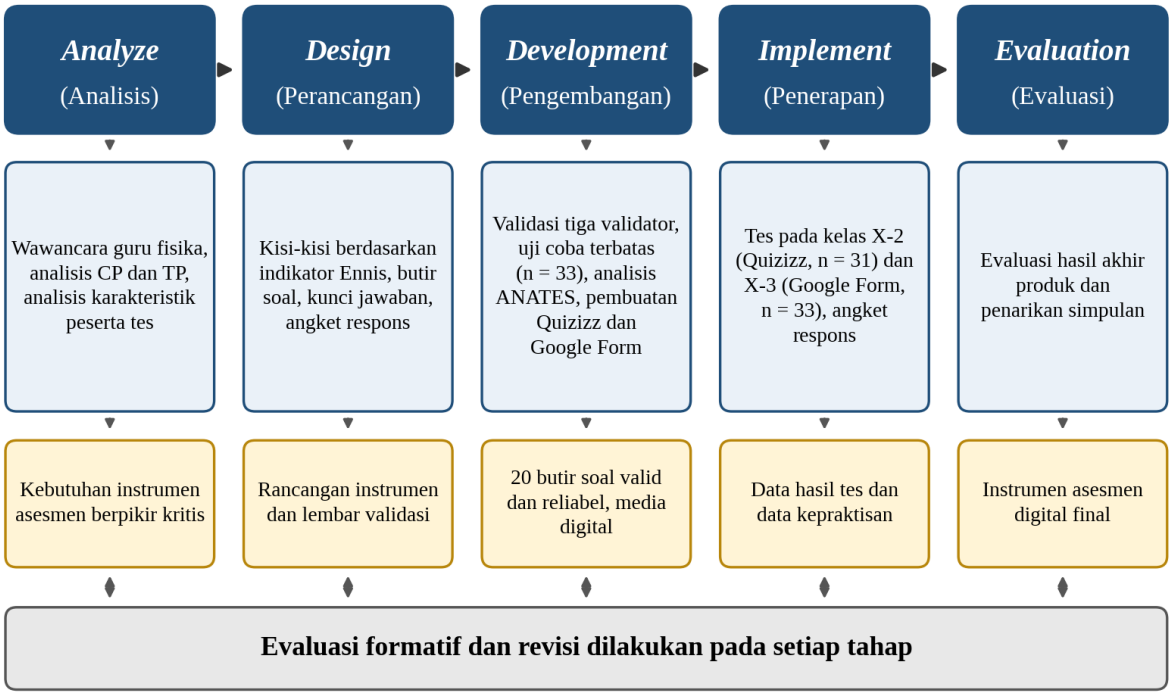
Berdasarkan uraian tersebut, terdapat celah penelitian yang jelas. Instrumen asesmen berpikir kritis pada materi energi terbarukan yang teruji secara teoretis dan empiris, disajikan pada dua *platform* digital yang banyak dipakai guru, serta dikaji kepraktisannya, belum tersedia. Kebaruan penelitian ini terletak pada pengembangan satu perangkat butir soal yang sama dan pengujiannya pada *Quizizz* maupun *Google Form*, sehingga guru dapat memilih *platform* sesuai kebutuhan tanpa kehilangan kualitas pengukuran. Kualitas instrumen ditelaah melalui validitas teoretis (disertai kesepakatan antar validator), validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda, sedangkan kepraktisan ditelaah melalui respons peserta didik. Rumusan masalah penelitian ini adalah: (1) bagaimana kualitas instrumen asesmen digital keterampilan berpikir kritis pada materi energi terbarukan yang dikembangkan, ditinjau dari validitas teoretis, validitas empiris, reliabilitas, tingkat kesukaran, dan daya pembeda; dan (2) bagaimana kepraktisan penggunaan *Quizizz* dan

Google Form sebagai media asesmen tersebut. Tujuan penelitian ini adalah menganalisis kualitas instrumen dan kepraktisan kedua *platform* tersebut.

METODE PENELITIAN

Desain Penelitian

Penelitian ini merupakan penelitian pengembangan (*Research and Development*), yaitu penelitian yang bertujuan menghasilkan produk tertentu dan menguji keefektifannya (Sugiyono, 2017). Model pengembangan yang digunakan adalah ADDIE yang diadaptasi dari Branch (2009), terdiri atas lima tahap, yaitu *Analyze* (analisis), *Design* (perancangan), *Development* (pengembangan), *Implement* (penerapan), dan *Evaluation* (evaluasi). Tahapan beserta kegiatan dan luaran pada penelitian ini disajikan pada Gambar 1. Butir soal disusun berdasarkan lima indikator berpikir kritis menurut Ennis (2011), yaitu memberikan penjelasan sederhana, membangun keterampilan dasar, menyimpulkan, memberikan penjelasan lanjut, serta mengatur strategi dan taktik, dengan tingkat kognitif analisis (C4) dan evaluasi (C5).



Gambar 1. Tahapan Pengembangan Instrumen Menggunakan Model ADDIE (diadaptasi dari Branch, 2009)

Prosedur Penelitian

Pada tahap *Analyze*, peneliti mewawancarai guru fisika SMA Katolik Karitas III Surabaya untuk mengetahui praktik asesmen yang berlangsung dan kebutuhan instrumen berpikir kritis, menelaah capaian pembelajaran (CP) dan tujuan pembelajaran (TP) materi energi terbarukan, serta menganalisis karakteristik peserta didik. Pada tahap *Design*, peneliti menyusun kisi-kisi berdasarkan indikator Ennis, butir soal pilihan ganda berstimulus, kunci jawaban beserta pembahasan, dan angket respons peserta didik. Pada tahap *Development*, instrumen divalidasi oleh tiga validator (dua dosen Program Studi Pendidikan Fisika dan satu guru fisika), direvisi sesuai masukan, diujicobakan secara terbatas kepada 33 peserta didik kelas X menggunakan tes cetak, dianalisis menggunakan aplikasi ANATES, kemudian disajikan pada *Quizizz* dan *Google Form*. Pada tahap *Implement*, paket soal yang sama diterapkan kepada peserta didik kelas X-2 melalui *Quizizz* dan kelas X-3 melalui *Google Form*, dilanjutkan pengisian angket respons sesuai *platform* yang digunakan. Tahap *Evaluation*

dilakukan pada setiap tahap untuk menyempurnakan kisi-kisi, butir soal, dan penyajian pada kedua *platform*.

Partisipan Penelitian

Partisipan terdiri atas tiga validator, 33 peserta didik kelas X pada uji coba terbatas, dan 64 peserta didik pada tahap penerapan, yaitu 31 peserta didik kelas X-2 (*Quizizz*) dan 33 peserta didik kelas X-3 (*Google Form*) SMA Katolik Karitas III Surabaya. Objek penelitian adalah instrumen asesmen keterampilan berpikir kritis pada materi energi terbarukan.

Instrumen dan Teknik Pengumpulan Data

Instrumen penelitian terdiri atas tiga jenis. Pertama, lembar validasi yang menilai aspek isi, konstruksi, dan bahasa dengan skala Likert 1–4 (Sugiyono, 2022). Kedua, tes keterampilan berpikir kritis berbentuk pilihan ganda dengan skor 1 untuk jawaban benar dan 0 untuk jawaban salah (Arikunto, 2018). Ketiga, angket respons peserta didik dengan skala Guttman, yaitu jawaban "Ya" bernilai 1 dan "Tidak" bernilai 0 (Sugiyono, 2017). Data validitas teoretis dikumpulkan melalui metode angket oleh validator, data empiris dan hasil belajar melalui metode tes, dan data kepraktisan melalui angket setelah peserta didik menggunakan *Quizizz* atau *Google Form*.

Teknik Analisis Data

Validitas teoretis dihitung sebagai persentase skor yang diperoleh terhadap skor maksimum menggunakan Persamaan 1.

$$P = \frac{skor\ yang\ diperoleh}{skor\ maksimum} \times 100\% \quad (1)$$

Hasilnya diinterpretasikan menurut Arifin (2016): 0–20% sangat kurang valid, 21–40% kurang valid, 41–60% cukup valid, 61–80% valid, dan 81–100% sangat valid. Instrumen dinyatakan layak apabila memperoleh persentase minimal 61%. Kesepakatan antar validator dihitung menggunakan *Percentage of Agreement* (PoA) pada Persamaan 2 (Borich, 1994), dengan A adalah skor yang lebih tinggi dan B skor yang lebih rendah dari sepasang validator. PoA dihitung pada setiap pasangan validator, kemudian dirata-ratakan, dan nilai PoA minimal 75% dipandang menunjukkan kesepakatan yang baik (Borich, 1994).

$$PoA = \left[1 - \frac{A-B}{A+B}\right] \times 100\% \quad (2)$$

Validitas empiris butir soal dianalisis menggunakan korelasi *Product Moment* Pearson (Persamaan 3) antara skor butir (X) dan skor total (Y), dengan N adalah banyaknya peserta uji coba (Arikunto, 2012). Butir dinyatakan valid apabila r hitung lebih besar daripada r tabel pada taraf signifikansi 5%.

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\Sigma X)^2][N\sum Y^2 - (\Sigma Y)^2]}} \quad (3)$$

Reliabilitas dihitung menggunakan rumus Kuder-Richardson 20 (Persamaan 4) karena tes berbentuk pilihan ganda dengan penskoran dikotomis dan tingkat kesukaran butir yang tidak seragam (Arikunto, 2012; Kuder & Richardson, 1937). Pada persamaan tersebut, k adalah banyaknya butir, S^2 adalah varians skor total, dan Σpq adalah jumlah hasil kali proporsi jawaban benar (p) dan salah (q) pada seluruh butir. Reliabilitas dikategorikan sangat tinggi (0,81–1,00), tinggi (0,61–0,80), sedang (0,41–0,60), rendah (0,21–0,40), dan sangat rendah (0,00–0,20), dan instrumen dinyatakan reliabel apabila $r_{11} \geq 0{,}41$ (Arikunto, 2012).

$$r_{11} = \frac{k}{k-1}\left(\frac{S^2 - \Sigma pq}{S^2}\right) \quad (4)$$

Tingkat kesukaran dihitung dengan Persamaan 5 (Arifin, 2016), dengan ΣX adalah jumlah skor seluruh peserta pada butir tersebut dan N adalah jumlah peserta. Kriterianya adalah sukar (0,00–0,30), sedang (0,31–0,70), dan mudah (0,71–1,00) (Arikunto, 2012). Daya pembeda dihitung dengan Persamaan 6 (Arikunto, 2018), dengan KA dan KB berturut-turut adalah rata-rata skor kelompok atas

dan kelompok bawah. Kategorinya adalah sangat jelek (DP ≤ 0,00), jelek (0,00–0,20), cukup (0,21–0,40), baik (0,41–0,70), dan sangat baik (0,71–1,00), dan butir dengan DP ≥ 0,21 dipertahankan.

$$TK = \frac{\frac{\sum X}{N}}{skor\ maksimum} \quad (5)$$

$$DP = \frac{K_A - K_B}{skor\ maksimum} \quad (6)$$

Data angket respons dianalisis secara deskriptif dengan persentase jawaban "Ya" pada setiap butir pernyataan (Persamaan 7), dengan F adalah jumlah peserta yang menjawab "Ya" dan N adalah jumlah peserta. Hasilnya dikategorikan menurut Riduwan (2015): 0–20% sangat kurang baik, 21–40% kurang baik, 41–60% cukup baik, 61–80% baik, dan 81–100% sangat baik. Media dinyatakan praktis apabila persentase respons minimal 61%. Skor tes hasil implementasi dianalisis secara deskriptif (nilai tertinggi, terendah, dan rata-rata) sebagai data pendukung keterlaksanaan, bukan untuk membandingkan efektivitas kedua *platform*.

$$P = \frac{F}{N} \times 100\% \quad (7)$$

HASIL DAN PEMBAHASAN

Hasil Penelitian

Tahap Analyze. Wawancara dengan guru fisika menunjukkan bahwa *Quizizz* dan *Google Form* telah digunakan dalam pembelajaran, tetapi belum dirancang untuk mengukur keterampilan berpikir kritis dan masih rentan terhadap kecurangan. Telaah CP dan TP menetapkan cakupan materi energi terbarukan (sumber energi surya, angin, air, biomassa, dan panas bumi) yang akan diukur. Hasil analisis ini menjadi dasar penyusunan instrumen pada tahap berikutnya.

Tahap Design. Peneliti menyusun kisi-kisi berdasarkan lima indikator berpikir kritis menurut Ennis (2011) pada tingkat kognitif C4 dan C5. Hasilnya adalah 20 butir soal pilihan ganda yang masing-masing dilengkapi stimulus (teks, gambar, grafik, atau tabel), kunci jawaban, dan pembahasan, serta angket respons peserta didik.

Tahap Development: validitas teoretis. Hasil validasi oleh tiga validator disajikan pada Tabel 1. Seluruh ranah berkategori sangat valid dengan rata-rata 93,75% (rata-rata seluruh 12 pernyataan). Kesepakatan antar validator tinggi, yaitu PoA 95,24% (pasangan validator 1–2 sebesar 97,62%, 1–3 sebesar 95,24%, dan 2–3 sebesar 92,86%), jauh di atas batas 75%. Validator menyarankan penambahan variasi stimulus berupa grafik dan tabel serta ilustrasi visual yang sesuai dengan fenomena pada soal. Saran tersebut ditindaklanjuti dengan merevisi beberapa butir sebelum uji coba terbatas.

Tabel 1. Hasil Validasi Ahli dan Kesepakatan Antar Validator pada Instrumen Tes

Ranah penilaian	Persentase validitas (%)	Kategori	PoA (%)
Isi	97,92	Sangat valid	97,62
Konstruksi	90,00	Sangat valid	94,29
Bahasa	94,44	Sangat valid	93,65
Rata-rata keseluruhan	**93,75**	**Sangat valid**	**95,24**

Catatan: persentase validitas dihitung dengan Persamaan 1 dari skor tiga validator pada 12 pernyataan (isi 4, konstruksi 5, bahasa 3); PoA dihitung dengan Persamaan 2 pada tiga pasangan validator. Rata-rata keseluruhan adalah rata-rata seluruh pernyataan.

Angket respons juga divalidasi oleh tiga validator dan berkategori sangat valid, dengan persentase validitas isi 88,89%, konstruksi 97,22%, dan bahasa 100%, serta PoA rata-rata 95,77%.

Tahap Development: validitas empiris. Uji coba terbatas pada 33 peserta didik menghasilkan data yang dianalisis menggunakan ANATES (Tabel 2). Seluruh 20 butir valid karena r hitung berada pada rentang 0,433–0,700 dan melampaui r tabel 0,344 (N = 33, taraf signifikansi 5%). Lima butir (nomor 6, 8, 11, 13, dan 15) berkategori sangat signifikan, sedangkan 15 butir lainnya berkategori signifikan. Koefisien reliabilitas sebesar 0,87 berkategori sangat tinggi. Tingkat kesukaran berkisar

0,15–0,70, dengan 18 butir berkategori sedang dan 2 butir (nomor 2 dan 16) berkategori sukar, tanpa butir yang berkategori mudah. Daya pembeda berkisar 0,44–0,89, dengan 16 butir berkategori baik dan 4 butir (nomor 5, 6, 8, dan 13) berkategori sangat baik.

Tabel 2. Hasil Analisis Empiris Butir Soal (N = 33)

No	r hitung	Kriteria	Keputusan	TK	Kategori TK	DP	Kategori DP
1	0,456	Signifikan	Valid	0,55	Sedang	0,56	Baik
2	0,460	Signifikan	Valid	0,15	Sukar	0,44	Baik
3	0,463	Signifikan	Valid	0,70	Sedang	0,56	Baik
4	0,433	Signifikan	Valid	0,67	Sedang	0,56	Baik
5	0,526	Signifikan	Valid	0,52	Sedang	0,78	Sangat baik
6	0,630	Sangat signifikan	Valid	0,70	Sedang	0,78	Sangat baik
7	0,487	Signifikan	Valid	0,67	Sedang	0,56	Baik
8	0,659	Sangat signifikan	Valid	0,64	Sedang	0,89	Sangat baik
9	0,436	Signifikan	Valid	0,70	Sedang	0,56	Baik
10	0,513	Signifikan	Valid	0,64	Sedang	0,67	Baik
11	0,585	Sangat signifikan	Valid	0,48	Sedang	0,67	Baik
12	0,433	Signifikan	Valid	0,33	Sedang	0,44	Baik
13	0,700	Sangat signifikan	Valid	0,70	Sedang	0,89	Sangat baik
14	0,487	Signifikan	Valid	0,67	Sedang	0,56	Baik
15	0,561	Sangat signifikan	Valid	0,70	Sedang	0,67	Baik
16	0,433	Signifikan	Valid	0,18	Sukar	0,44	Baik
17	0,450	Signifikan	Valid	0,70	Sedang	0,56	Baik
18	0,435	Signifikan	Valid	0,61	Sedang	0,56	Baik
19	0,528	Signifikan	Valid	0,67	Sedang	0,56	Baik
20	0,547	Signifikan	Valid	0,70	Sedang	0,67	Baik

Catatan: r tabel = 0,344 (taraf signifikansi 5%); TK = tingkat kesukaran; DP = daya pembeda. Nilai TK dan DP dinyatakan dalam indeks (persentase dibagi 100) agar sesuai dengan kriteria pada bagian metode.

Tahap Implement. Instrumen yang sama disajikan pada *Quizizz* (kelas X-2, 31 peserta didik) dan *Google Form* (kelas X-3, 33 peserta didik). Hasil angket respons (Tabel 3) menunjukkan bahwa kedua *platform* berkategori sangat praktis, yaitu 94,98% untuk *Quizizz* dan 94,95% untuk *Google Form*. Pada *Quizizz*, seluruh pernyataan berkategori sangat praktis. Pada *Google Form*, hanya pernyataan tentang kenyamanan penggunaan yang berkategori praktis (78,78%), sedangkan pernyataan lainnya berkategori sangat praktis (87,87–100%).

Tabel 3. Hasil Angket Respons Peserta Didik terhadap Kepraktisan Media

Media	Rata-rata (%)	Pernyataan terendah	Kategori
Quizizz	94,98	Seluruh pernyataan sangat praktis	Sangat praktis
Google Form	94,95	Kenyamanan penggunaan (78,78%)	Sangat praktis

Statistik deskriptif hasil tes (Tabel 4) menunjukkan seluruh peserta didik pada kedua kelas dapat menyelesaikan instrumen melalui *platform* masing-masing. Data ini digunakan sebagai pendukung keterlaksanaan, bukan untuk menyimpulkan keunggulan salah satu *platform*, karena kedua kelas dapat berbeda kemampuan awalnya.

Tabel 4. Statistik Deskriptif Hasil Tes pada Tahap Implement

Media	Jumlah peserta	Nilai tertinggi	Nilai terendah	Rata-rata
Quizizz (X-2)	31	95	50	73,06
Google Form (X-3)	33	95	10	60,30

Tahap Evaluation. Evaluasi pada setiap tahap menghasilkan perbaikan penyajian stimulus dan visualisasi pada tahap *Design*, serta pengecekan kesetaraan isi pada kedua *platform* pada tahap *Development*. Pada tahap *Implement*, ditemukan tiga kendala: petunjuk pengerjaan *Quizizz* perlu dijelaskan secara lisan dan berulang; beberapa peserta didik keluar dari sesi *Quizizz* karena gangguan koneksi sehingga harus mengerjakan ulang dari awal; dan *Quizizz* tidak menyediakan kolom identitas sehingga pencocokan data dilakukan manual berdasarkan nama pengguna.

Pembahasan

Kualitas instrumen. Validitas teoretis yang sangat valid pada ketiga ranah (isi 97,92%, konstruksi 90,00%, bahasa 94,44%) menunjukkan bahwa butir soal sesuai dengan CP, TP, dan indikator berpikir kritis Ennis, tersusun sesuai kaidah soal pilihan ganda, dan menggunakan bahasa yang jelas. Nilai PoA 95,24% memperkuat temuan ini karena menunjukkan bahwa penilaian tidak ditentukan oleh satu validator saja. Ranah konstruksi menjadi yang terendah, antara lain karena pernyataan P6 pada lembar validasi memperoleh skor 3 dari ketiga validator (75%). Masukan validator tentang variasi stimulus dan ilustrasi telah ditindaklanjuti sebelum uji coba terbatas. Capaian ini sejalan dengan Karira dkk. (2023) yang memperoleh validitas teoretis 98% pada instrumen berpikir kritis materi energi terbarukan dengan tiga validator, serta Alifah dan Sunarti (2024) yang menyatakan instrumen berbasis *Google Form* sangat valid menurut tiga ahli. Safitri dkk. (2023) melaporkan validitas 86,19% pada instrumen berbasis *Quizizz*, sedikit di bawah capaian penelitian ini, yang dapat dikaitkan dengan perbedaan jenjang, materi, dan bentuk soal.

Pada validitas empiris, seluruh butir valid (r hitung 0,433–0,700 > 0,344). Temuan ini berbeda dengan Karira dkk. (2023) yang hanya memperoleh sembilan butir valid dari 15 butir meskipun validitas teoretisnya 98%, dan mereka menyimpulkan bahwa butir yang valid menurut ahli belum tentu valid ketika diujikan kepada peserta didik. Perbedaan hasil dapat dijelaskan oleh beberapa faktor. Pertama, penelitian ini menggunakan pilihan ganda berstimulus dengan penskoran 0–1 yang menghasilkan korelasi butir-total yang lebih stabil, sedangkan Karira dkk. memakai soal uraian dengan rubrik bertingkat. Kedua, revisi stimulus berdasarkan masukan validator dilakukan sebelum uji coba, sehingga kualitas butir sudah diperbaiki lebih dahulu. Ketiga, ukuran sampel berbeda (33 dan 30 peserta didik) dan karakteristik sekolah tidak sama. Penjelasan ini bersifat dugaan karena kedua penelitian tidak dirancang untuk dibandingkan secara langsung. Reliabilitas 0,87 (sangat tinggi) menunjukkan konsistensi pengukuran yang baik. Nilai yang sama dilaporkan oleh Karira dkk. (2023), sedangkan Safitri dkk. (2023) melaporkan 0,791 untuk soal pilihan ganda dan isian singkat. Menurut Arikunto (2012), reliabilitas menunjukkan konsistensi hasil pengukuran, sehingga nilai ini menunjukkan bahwa instrumen layak dipakai untuk pengukuran berulang.

Tingkat kesukaran didominasi kategori sedang (18 butir), dengan dua butir sukar (nomor 2 dan 16) dan tanpa butir mudah. Komposisi ini lebih seimbang daripada yang dilaporkan Safitri dkk. (2023), yang 50% butirnya berkategori mudah. Dua butir sukar tetap dipertahankan karena daya pembedanya baik (0,44) dan berfungsi mengidentifikasi peserta didik dengan kemampuan berpikir kritis yang lebih tinggi; butir nomor 16 menuntut peserta didik mengenali asumsi tersembunyi, yaitu indikator yang relatif menuntut penalaran lebih lanjut. Seluruh butir memiliki daya pembeda baik hingga sangat baik (0,44–0,89), lebih baik daripada Safitri dkk. (2023) yang masih memperoleh empat butir berdaya pembeda jelek. Kualitas ini sesuai dengan pendapat Arikunto (2018) bahwa butir yang baik dapat membedakan peserta didik berkemampuan tinggi dan rendah.

Kepraktisan. Kedua *platform* berkategori sangat praktis (94,98% dan 94,95%) dengan selisih yang sangat kecil, sehingga keduanya dapat dipakai sebagai media penyaji instrumen yang sama. Hasil ini sejalan dengan Safitri dkk. (2023) yang melaporkan respons peserta didik terhadap instrumen berbasis *Quizizz* berkategori sangat baik (88,9%) dan Azizah dan Hidayat (2024) yang menyatakan asesmen digital berbasis *Quizizz* praktis. Perbedaan kecil muncul pada pernyataan kenyamanan *Google Form* (78,78%). Tampilan *Google Form* yang lebih formal dan minim unsur permainan dapat menjadi penyebab kenyamanan yang sedikit lebih rendah, meskipun penelitian ini tidak menguji penyebab tersebut secara langsung.

Kendala pada *Quizizz* sejalan dengan Safitri dkk. (2023) yang menemukan sebagian peserta didik masih bingung menggunakan *Quizizz* sehingga petunjuk perlu dijelaskan ulang. Hal ini menunjukkan bahwa kepraktisan *Quizizz* sangat bergantung pada pengenalan fitur sebelum tes dan stabilitas koneksi. Pada *Google Form*, pengisian identitas lebih mudah karena formulir identitas tersedia sehingga pengolahan data lebih efisien. Perbedaan rata-rata skor antara kelas *Quizizz* (73,06) dan *Google Form* (60,30) tidak dapat ditafsirkan sebagai perbedaan efektivitas *platform* karena kemampuan awal kedua kelas tidak dikendalikan.

Implikasi. Bagi guru, penelitian ini menyediakan 20 butir soal berpikir kritis materi energi terbarukan yang telah teruji dan dapat dipakai pada *Quizizz* maupun *Google Form*. Pemilihan *platform* dapat disesuaikan dengan kondisi: *Quizizz* cocok bila guru menginginkan suasana yang interaktif dan koneksi stabil, sedangkan *Google Form* lebih mudah untuk pencatatan identitas dan pengolahan data. Bagi sekolah, instrumen ini dapat menjadi contoh asesmen yang mendukung penguatan Profil Pelajar Pancasila pada dimensi bernalar kritis. Bagi peneliti, prosedur yang memadukan validasi ahli dengan kesepakatan antar validator, uji empiris, dan kajian kepraktisan pada dua *platform* dapat diterapkan pada materi fisika lain.

Keterbatasan. Instrumen hanya mencakup materi energi terbarukan, subjek hanya dua kelas di satu sekolah, dan hanya dua *platform* yang diuji. Selain itu, bentuk pilihan ganda belum menangkap seluruh proses berpikir kritis. Karira dkk. (2023) merujuk Ennis (2011) bahwa soal uraian lebih tepat untuk mengukur berpikir kritis karena pilihan ganda berisiko kehilangan indikator yang diukur. Penelitian selanjutnya dapat mengembangkan butir uraian atau pilihan ganda beralasan, melibatkan sampel yang lebih besar, dan menguji efektivitas kedua *platform* dengan desain yang mengendalikan kemampuan awal.

SIMPULAN

Instrumen asesmen digital keterampilan berpikir kritis pada materi energi terbarukan yang dikembangkan dengan model ADDIE berkualitas baik. Secara teoretis, instrumen berkategori sangat valid (93,75%) dengan kesepakatan antar validator yang tinggi (PoA 95,24%). Secara empiris, seluruh 20 butir valid (r hitung 0,433–0,700), reliabilitas sebesar 0,87 (sangat tinggi), tingkat kesukaran didominasi kategori sedang (18 butir) dengan dua butir sukar, dan daya pembeda berkategori baik hingga sangat baik (0,44–0,89). Penggunaan *Quizizz* dan *Google Form* sebagai media asesmen berkategori sangat praktis, masing-masing sebesar 94,98% dan 94,95%. Dengan demikian, instrumen ini layak digunakan guru sebagai alternatif asesmen keterampilan berpikir kritis pada materi energi terbarukan melalui salah satu dari kedua *platform*. Saran untuk penelitian selanjutnya adalah memperluas cakupan materi dan sampel, menambahkan bentuk soal yang menuntut alasan, serta menguji efektivitas kedua *platform* dengan desain penelitian yang sesuai.

KONTRIBUSI PENULIS

Graciella Setiarini: konseptualisasi, metodologi, investigasi, analisis formal, kurasi data, dan penulisan draf awal. Abu Zainuddin: konseptualisasi, metodologi, validasi, supervisi, serta penulisan, peninjauan, dan penyuntingan. Semua penulis telah membaca dan menyetujui versi akhir naskah ini.

PERNYATAAN BEBAS KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa tidak ada konflik kepentingan finansial maupun hubungan pribadi yang dapat memengaruhi hasil yang dilaporkan dalam naskah ini.

PERNYATAAN ETIKA PENELITIAN DAN PUBLIKASI

Para penulis menyatakan bahwa penelitian dan penulisan naskah ini telah mematuhi standar etika penelitian dan publikasi, sesuai dengan prinsip ilmiah, serta bebas dari plagiasi.

PERNYATAAN PEMANFAATAN TEKNOLOGI ASISTIF

Para penulis menyatakan bahwa kecerdasan buatan generatif (*Generative Artificial Intelligence*) dan teknologi asistif lainnya digunakan secara terbatas dalam penulisan naskah ini untuk membantu penyusunan struktur naskah, penyempurnaan tata bahasa, dan penyuntingan kalimat. [Lengkapi nama alat yang benar-benar digunakan.] Para penulis telah meninjau, memverifikasi, dan menyunting seluruh konten yang dihasilkan guna memastikan ketepatan, kesesuaian dengan data penelitian, serta kepatuhan terhadap standar etika dan ilmiah, dan bertanggung jawab penuh atas naskah versi akhir.

DAFTAR PUSTAKA

Alifah, D. R. Z. Z., & Sunarti, T. (2024). Pengembangan instrumen tes keterampilan berpikir kritis berbasis G-Form pada materi Asas Black. *Inovasi Pendidikan Fisika, 13*(3), 222–230. https://ejournal.unesa.ac.id/index.php/inovasi-pendidikan-fisika/article/view/62628

Arifin, Z. (2016). *Evaluasi pembelajaran*. Bandung: PT Remaja Rosdakarya.

Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan* (Edisi 2). Jakarta: Bumi Aksara.

Arikunto, S. (2018). *Dasar-dasar evaluasi pendidikan* (Edisi 3). Jakarta: Bumi Aksara.

Azizah, N. N., & Hidayat, R. (2024). Pengembangan asesmen digital berbasis game edukatif Quizizz pada mata pelajaran marketing. *Journal of Innovation and Teacher Professionalism, 2*(2), 195–209.

Balqis, A. S., & Andriani, A. E. (2024). Development of learning evaluation based on automatic assessment through Quizizz Paper Mode to improve students' natural and social sciences learning outcomes. *Jurnal Penelitian Pendidikan IPA, 10*(6), 3357–3366.

Borich, G. D. (1994). *Observation skills for effective teaching*. New York: Macmillan Publishing Company.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. New York: Springer.

Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions and abilities*. University of Illinois.

Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Hermosa Beach, CA: Measured Reasons LLC.

Karira, N. F., Sunarti, T., Jauhariyah, M. N. R., & Setyarsih, W. (2023). Validitas instrumen tes berbasis literasi sains untuk mengukur keterampilan berpikir kritis siswa SMA pada materi energi terbarukan. *Inovasi Pendidikan Fisika, 12*(2), 118–125. https://ejournal.unesa.ac.id/index.php/inovasi-pendidikan-fisika/article/view/52957

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151–160. https://doi.org/10.1007/BF02288391

Riduwan. (2015). *Skala pengukuran variabel-variabel penelitian*. Bandung: Alfabeta.

Safitri, E. M., Wahyuni, S., & Ahmad, N. (2023). Pengembangan instrumen asesmen keterampilan berpikir kritis menggunakan aplikasi Quizizz pada mata pelajaran IPA SMP. *Quantum: Jurnal Inovasi Pendidikan Sains, 14*(1), 96–106.

Sugiyono. (2017). *Metode penelitian dan pengembangan*. Bandung: Alfabeta.

Sugiyono. (2022). *Metode penelitian kuantitatif*. Bandung: Alfabeta.