

---

## Analisis Butir Soal *Multiple Choice* Materi Posisi Strategis Indonesia dan Potensi SDA Kelas XI SMA

Aulia Nailatu Az-zakia Rahmani<sup>1</sup>, Rizka Rahmalia<sup>2</sup>, Windy Ariyanti<sup>3</sup>, Mamat Ruhimat<sup>4</sup>, Dina Siti Logayah<sup>5</sup>, Igna Juli Triana<sup>6</sup>

Geography Education, Indonesian Education University, Bandung, Indonesia<sup>1-6</sup>

Email Korespondensi: [aulianaila@upi.edu](mailto:aulianaila@upi.edu), [rizkarhmlia25@upi.edu](mailto:rizkarhmlia25@upi.edu), [windyariyanti@upi.edu](mailto:windyariyanti@upi.edu), [mamatruhimat@upi.edu](mailto:mamatruhimat@upi.edu), [dina.logayah@upi.edu](mailto:dina.logayah@upi.edu), [ignajulitriana@upi.edu](mailto:ignajulitriana@upi.edu)

---

Article received: 27 Juni 2025, Review process: 13 Juli 2025,

Article Accepted: 25 Agustus 2025, Article published: 31 Agustus 2025

---

### ABSTRACT

The 21st-century education paradigm demands a transformation toward developing Higher Order Thinking Skills (HOTS) to enhance students' analytical, evaluative, and creative abilities. This challenge requires educators to design high-quality assessment instruments aligned with international standards. This study aims to analyze the quality of multiple-choice test items on the topic of Indonesia's strategic position and natural resource potential for Grade XI senior high school students. A quantitative descriptive approach was applied with a sample of 38 students selected through a random sampling technique. The research instrument consisted of 30 HOTS-based geography items, evaluated for validity, reliability, difficulty level, discrimination index, and distractor effectiveness. The findings revealed that 27 items (90%) were valid, the instrument demonstrated high reliability ( $r = 0.913$ ), the difficulty level was dominated by moderate items (70%), discrimination indices indicated 66% moderate, 30% good, and 3% strong items, while 81.33% of distractors were effective.

**Keywords:** item analysis, HOTS, validity, reliability, competency-based assessment

### ABSTRAK

Pendidikan abad ke-21 menuntut adanya transformasi paradigma pembelajaran yang berorientasi pada pengembangan Higher Order Thinking Skills (HOTS) untuk meningkatkan kemampuan analisis, evaluasi, dan sintesis peserta didik. Tantangan ini menuntut pendidik merancang instrumen penilaian yang berkualitas dan sesuai dengan standar internasional. Penelitian ini bertujuan untuk menganalisis kualitas butir soal *multiple choice* pada materi posisi strategis Indonesia dan potensi sumber daya alam untuk siswa kelas XI SMA. Pendekatan penelitian menggunakan kuantitatif deskriptif dengan sampel 38 siswa yang dipilih melalui teknik *random sampling*. Instrumen penelitian berupa 30 butir soal geografi berbasis HOTS yang diuji menggunakan analisis validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil penelitian menunjukkan bahwa 27 butir soal (90%) dinyatakan valid, reliabilitas instrumen tinggi ( $r = 0,913$ ), tingkat kesukaran didominasi kategori sedang (70%), daya pembeda soal berada pada kategori sedang (66%), baik (30%), dan kuat (3%), serta 81,33% opsi jawaban efektif.

**Kata Kunci:** analisis butir soal, HOTS, validitas, reliabilitas, penilaian berbasis kompetensi

## PENDAHULUAN

Pendidikan abad ke-21 menuntut adanya transformasi paradigma dalam proses pembelajaran dan penilaian hasil belajar. Tantangan utama pendidikan saat ini terletak pada upaya mengembangkan keterampilan berpikir tingkat tinggi atau *Higher Order Thinking Skills* (HOTS) yang berperan dalam membentuk kemampuan analisis, evaluasi, dan sintesis peserta didik. Anderson dan Krathwohl (2001) menegaskan bahwa HOTS mencakup level kognitif C4 (analisis), C5 (evaluasi), dan C6 (kreasi), yang menuntut peran pendidik dalam merancang soal yang memicu kemampuan berpikir kritis, logis, dan kreatif. Studi internasional oleh Brookhart (2010) menunjukkan bahwa integrasi HOTS dalam pembelajaran meningkatkan kualitas hasil belajar sekaligus mempersiapkan peserta didik menghadapi dinamika global yang semakin kompetitif.

Permasalahan mendasar yang muncul di tingkat implementasi adalah kesalahpahaman pendidik mengenai konsep HOTS. Sebagian besar guru masih mengasosiasikan HOTS dengan soal yang sulit dan rumit tanpa mempertimbangkan esensi keterampilan berpikir tingkat tinggi itu sendiri. Hasil penelitian OECD (2019) menegaskan bahwa soal berkualitas tidak hanya menguji daya ingat, tetapi juga mengukur kemampuan siswa dalam mengaplikasikan konsep pada situasi nyata. Penerapan HOTS menuntut pembiasaan berpikir reflektif dan analitis, bukan sekadar memberikan soal yang kompleks. Oleh karena itu, pemahaman guru mengenai prinsip konstruksi soal HOTS menjadi aspek krusial yang perlu diperkuat.

Penilaian hasil belajar pada kurikulum abad ke-21 menekankan pentingnya keseimbangan antara aspek kognitif, afektif, dan psikomotorik. Instrumen evaluasi yang berkualitas harus mampu mengukur kemampuan berpikir kritis, kolaboratif, komunikatif, dan kreatif (*4C Skills*). Menurut Wiliam (2018), instrumen yang reliabel tidak hanya memastikan hasil penilaian yang konsisten, tetapi juga relevan dengan kebutuhan pengembangan keterampilan abad ke-21. Analisis kualitas butir soal menjadi bagian penting dalam memastikan bahwa instrumen penilaian tidak hanya akurat, tetapi juga berfungsi sebagai alat diagnostik yang membantu guru memahami perkembangan kompetensi peserta didik.

Tantangan implementasi instrumen berbasis HOTS semakin kompleks dengan adanya penerapan Kurikulum 2013 yang menekankan penilaian autentik dan berbasis kompetensi. Realitas di lapangan menunjukkan bahwa sebagian guru masih mengalami kendala dalam merancang indikator, kisi-kisi, dan butir soal yang sesuai standar internasional. Penelitian Darling-Hammond et al. (2020) mengonfirmasi bahwa hambatan ini disebabkan oleh kurangnya pelatihan teknis dan pemahaman konseptual. Instrumen penilaian yang disusun tanpa mempertimbangkan validitas, reliabilitas, tingkat kesukaran, dan daya pembeda akan menghasilkan evaluasi yang bias dan tidak sesuai dengan tujuan pendidikan.

Perkembangan penelitian internasional mengenai asesmen berbasis HOTS memberikan bukti bahwa instrumen berkualitas mampu meningkatkan efektivitas pembelajaran dan memperkuat pencapaian kompetensi siswa. Menurut Marzano dan Heflebower (2016), asesmen yang dirancang berbasis HOTS mendorong siswa

untuk membangun koneksi antar-konsep dan mengembangkan kemampuan problem-solving secara mendalam. Temuan serupa juga diungkapkan oleh OECD (2021) melalui laporan *PISA*, yang menunjukkan bahwa integrasi soal HOTS dalam pembelajaran berdampak signifikan terhadap kesiapan siswa menghadapi tantangan global. Dengan demikian, asesmen yang tepat bukan hanya alat ukur, tetapi juga sarana pembelajaran yang memotivasi peserta didik.

Penelitian ini bertujuan untuk menganalisis kualitas butir soal *multiple choice* pada materi posisi strategis Indonesia dan potensi sumber daya alam untuk siswa kelas XI SMA. Analisis dilakukan berdasarkan parameter validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pengembangan instrumen evaluasi berbasis HOTS yang sesuai dengan standar internasional, relevan dengan tuntutan pendidikan abad ke-21, dan dapat menjadi acuan bagi guru dalam merancang penilaian yang lebih efektif.

## METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan tujuan menganalisis kualitas butir soal *multiple choice* pada materi posisi strategis Indonesia dan potensi sumber daya alam untuk siswa kelas XI SMA. Sampel penelitian berjumlah 38 siswa yang dipilih melalui teknik *random sampling* untuk memastikan representativitas populasi. Instrumen penelitian berupa 30 butir soal geografi yang dikembangkan berbasis *Higher Order Thinking Skills* (HOTS) sesuai taksonomi Bloom revisi. Proses pengumpulan data dilakukan melalui tes daring menggunakan platform *Quizizz* dan dianalisis menggunakan perangkat lunak Microsoft Excel untuk menghitung validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh. Validitas soal diuji menggunakan rumus Pearson Product Moment, sedangkan reliabilitas diukur dengan Kuder Richardson (KR-20) untuk memastikan konsistensi internal instrumen. Analisis tingkat kesukaran dan daya pembeda dilakukan dengan membandingkan proporsi jawaban benar antara kelompok atas dan kelompok bawah, sementara efektivitas pengecoh ditentukan berdasarkan distribusi pilihan jawaban responden. Pendekatan ini dipilih karena metode analisis kuantitatif deskriptif dinilai efektif dalam mengevaluasi kualitas instrumen pembelajaran dan memastikan kesesuaian soal dengan standar penilaian berbasis kompetensi abad ke-21 (Creswell & Creswell, 2018; Brookhart, 2010).

## HASIL DAN PEMBAHASAN

Hasil analisis terhadap 30 butir soal geografi yang diuji pada siswa kelas XI SMA menunjukkan bahwa sebagian besar soal telah memenuhi kriteria kualitas instrumen evaluasi yang baik dan dapat digunakan sebagai alat ukur kompetensi siswa. Uji validitas menggunakan metode Pearson Product Moment menghasilkan temuan bahwa 27 butir soal (90%) dinyatakan valid, sedangkan 3 butir soal (10%) dikategorikan tidak valid. Tingginya persentase soal valid menunjukkan bahwa proses penyusunan instrumen dilakukan secara sistematis, dimulai dari analisis

kompetensi dasar, penyusunan indikator pembelajaran, hingga penyesuaian butir soal dengan kerangka HOTS sesuai taksonomi Bloom revisi. Menurut *Brookhart* (2010), validitas yang tinggi pada instrumen tes mencerminkan kesesuaian antara tujuan pembelajaran dan indikator yang diukur, sehingga hasil evaluasi lebih representatif terhadap kemampuan siswa secara menyeluruh. Kondisi ini menunjukkan bahwa instrumen penilaian yang dikembangkan dalam penelitian ini telah memenuhi kriteria yang sesuai dengan standar asesmen berbasis kompetensi abad ke-21 dan dapat digunakan sebagai acuan oleh pendidik untuk mengukur pencapaian hasil belajar siswa secara lebih akurat.

**Tabel 1. Distribusi Validitas Butir Soal**

| Kategori     | Nomor Soal              | Jumlah      | Persentase |
|--------------|-------------------------|-------------|------------|
| Valid        | 1-16, 18, 19, 21, 23-30 | 27          | 90%        |
| Tidak Valid  | 17, 20, 22              | 3           | 10%        |
| <b>Total</b> | <b>30</b>               | <b>100%</b> |            |

Uji reliabilitas instrumen dilakukan menggunakan metode Kuder-Richardson (KR-20), dan hasilnya menunjukkan nilai koefisien sebesar 0,913, yang dikategorikan sangat tinggi. Nilai ini menegaskan bahwa instrumen memiliki konsistensi internal yang baik, artinya jika soal diujikan kembali pada kelompok responden yang berbeda dengan kondisi yang sama, maka hasil pengukurannya akan relatif stabil dan tidak mengalami perbedaan signifikan. Temuan ini sejalan dengan standar internasional yang menyatakan bahwa instrumen asesmen dikatakan reliabel apabila memiliki koefisien di atas 0,70 (*Cohen et al., 2018*). Dalam konteks penelitian ini, reliabilitas yang tinggi dapat diatribusikan pada proses penyusunan soal yang terstruktur, penentuan indikator pembelajaran yang tepat, serta pemilihan tingkat kesulitan yang proporsional dengan kompetensi siswa. Selain itu, keberhasilan menjaga konsistensi kualitas soal juga mencerminkan pentingnya proses telaah ahli dan validasi instrumen sebelum digunakan pada pengukuran skala besar, sebagaimana dianjurkan oleh *William* (2018) dalam penilaian berbasis HOTS.

Hasil analisis tingkat kesukaran menunjukkan bahwa 9 soal (30%) berada pada kategori mudah, 21 soal (70%) termasuk kategori sedang, dan tidak ada soal yang tergolong sukar. Distribusi ini menunjukkan bahwa sebagian besar butir soal telah disusun pada tingkat kesukaran yang ideal dan sesuai dengan kriteria penilaian berbasis kompetensi. Soal yang berada pada kategori sedang merupakan tipe soal yang memberikan tantangan intelektual yang seimbang, tidak terlalu sulit sehingga membuat siswa frustrasi, dan tidak terlalu mudah sehingga gagal menstimulasi keterampilan berpikir kritis. Menurut *Anderson & Krathwohl* (2001), proporsi soal kategori sedang paling tepat digunakan untuk mengukur kemampuan berpikir tingkat tinggi karena mendorong siswa untuk menghubungkan konsep-konsep dan menemukan pola baru. Dengan demikian, distribusi tingkat kesukaran dalam penelitian ini menunjukkan kesesuaian

instrumen dengan prinsip-prinsip HOTS yang menjadi fokus utama pengembangan soal geografi di era pendidikan modern.

*Tabel 2. Distribusi Tingkat Kesukaran Soal*

| Kategori     | Nomor Soal                         | Jumlah      | Persentase |
|--------------|------------------------------------|-------------|------------|
| Mudah        | 16, 18, 21, 22, 25, 26, 28, 29, 30 | 9           | 30%        |
| Sedang       | 1-15, 17, 19, 20, 23, 24, 27       | 21          | 70%        |
| Sukar        | -                                  | 0           | 0%         |
| <b>Total</b> | <b>30</b>                          | <b>100%</b> |            |

Daya pembeda soal diukur untuk mengetahui sejauh mana butir soal mampu membedakan siswa dengan kemampuan tinggi dan rendah. Hasil analisis menunjukkan bahwa 20 soal (66%) berada pada kategori sedang, 9 soal (30%) berkategori baik, dan 1 soal (3%) termasuk kategori sangat baik. Tidak ditemukan soal dengan daya pembeda rendah ataupun negatif, sehingga dapat disimpulkan bahwa sebagian besar soal efektif dalam mengklasifikasikan siswa berdasarkan tingkat penguasaan materi. Menurut *Haladyna & Rodriguez* (2013), instrumen dengan daya pembeda tinggi mampu memberikan gambaran yang lebih jelas tentang keberhasilan proses pembelajaran, karena dapat mengidentifikasi siswa dengan pemahaman konsep yang mendalam dibandingkan siswa yang hanya mengandalkan hafalan. Oleh karena itu, butir soal yang dikembangkan dalam penelitian ini telah dirancang dengan kualitas yang memadai, baik dari segi redaksi, pilihan jawaban, maupun konteks materi yang relevan, sehingga instrumen dapat dijadikan acuan dalam penyusunan soal yang lebih adaptif terhadap kebutuhan pembelajaran berbasis kompetensi.

Efektivitas pengecoh merupakan salah satu indikator penting dalam mengukur kualitas instrumen berbasis pilihan ganda. Dari total 150 opsi jawaban, hasil analisis menunjukkan bahwa 122 opsi (81,33%) dinyatakan efektif, 20 opsi (13,33%) perlu direvisi, dan 8 opsi (5,33%) ditolak karena dianggap tidak berfungsi optimal. Pengecoh yang tidak efektif dapat mengurangi kualitas soal karena memberikan peluang lebih besar bagi siswa untuk menebak jawaban benar tanpa memahami materi secara mendalam. Menurut *Haladyna, Downing, & Rodriguez* (2010), pengecoh yang efektif harus dirancang berdasarkan kesalahan konsep umum yang sering dilakukan siswa, sehingga dapat menilai tingkat pemahaman yang sebenarnya. Dengan demikian, perbaikan terhadap pengecoh yang kurang efektif perlu dilakukan agar instrumen lebih valid dan dapat mengukur kompetensi siswa secara akurat.

Secara keseluruhan, hasil analisis menunjukkan bahwa instrumen soal geografi yang dikembangkan memiliki kualitas tinggi dan dapat diandalkan untuk mengevaluasi kompetensi siswa pada materi posisi strategis Indonesia dan potensi sumber daya alam. Tingginya tingkat validitas dan reliabilitas serta proporsi tingkat kesukaran dan daya pembeda yang seimbang menunjukkan bahwa

instrumen telah memenuhi kriteria evaluasi internasional. Penelitian sebelumnya oleh *Cohen et al.* (2018) juga menegaskan bahwa kualitas instrumen yang baik akan mempengaruhi akurasi hasil evaluasi dan menjadi dasar dalam menyusun strategi pembelajaran yang lebih efektif. Oleh karena itu, instrumen ini dapat digunakan sebagai referensi untuk pengembangan asesmen berbasis HOTS dalam kurikulum abad ke-21.

Hasil analisis instrumen memberikan implikasi signifikan bagi guru dan pengembang kurikulum dalam meningkatkan kualitas penilaian hasil belajar siswa. Instrumen yang memenuhi kriteria validitas dan reliabilitas tinggi dapat digunakan untuk mengidentifikasi tingkat capaian kompetensi siswa sekaligus mengevaluasi efektivitas metode pembelajaran yang diterapkan. Temuan penelitian ini juga memperkuat hasil studi *OECD (2021)* yang menegaskan bahwa asesmen berbasis HOTS memegang peranan penting dalam mempersiapkan siswa menghadapi tantangan global, terutama dalam mengembangkan keterampilan analitis, kreatif, dan problem solving. Dengan demikian, instrumen penilaian ini dapat dijadikan acuan untuk perbaikan kurikulum dan penyusunan soal berbasis kompetensi di tingkat SMA secara lebih terarah dan berkelanjutan.

## Pembahasan

Hasil penelitian menunjukkan bahwa sebagian besar butir soal telah memenuhi indikator HOTS, terutama pada aspek analisis, evaluasi, dan kreasi. Instrumen yang digunakan menuntut siswa untuk menganalisis informasi, mengevaluasi argumen, dan mengembangkan jawaban berbasis penalaran tingkat tinggi. Studi internasional oleh *Brookhart (2010)* menegaskan bahwa penerapan HOTS dalam asesmen meningkatkan keterampilan berpikir kritis, logis, dan reflektif pada peserta didik. Integrasi instrumen berbasis HOTS dalam penelitian ini sejalan dengan tren global pendidikan yang menekankan pembelajaran berbasis kompetensi untuk meningkatkan kemampuan adaptasi dan literasi abad ke-21.

Hasil penelitian memperlihatkan kesesuaian instrumen dengan pendekatan kurikulum abad ke-21 yang berorientasi pada pengembangan keterampilan berpikir tingkat tinggi. Temuan ini mendukung kajian *Voogt & Roblin (2012)* yang menekankan bahwa instrumen asesmen perlu dirancang untuk mengukur keterampilan berpikir analitis dan kreatif dalam konteks permasalahan nyata. Dengan demikian, instrumen berbasis HOTS seperti dalam penelitian ini menjadi relevan dalam mempersiapkan siswa menghadapi tantangan global, sekaligus mendorong peningkatan kualitas pembelajaran yang lebih berorientasi pada keterampilan kompleks.

Tingkat validitas dan reliabilitas instrumen yang tinggi menegaskan peran penting pengembangan asesmen berbasis data dan pengukuran terstandar. Nilai koefisien reliabilitas sebesar 0,913 menunjukkan stabilitas hasil penilaian dan mendukung pandangan *Cohen et al.* (2018) bahwa instrumen yang reliabel mampu memberikan hasil evaluasi yang konsisten. Temuan ini selaras dengan standar internasional dan praktik terbaik dalam pengembangan instrumen pembelajaran

berbasis HOTS, di mana kualitas asesmen menjadi faktor kunci keberhasilan implementasi kurikulum berbasis kompetensi.

Distribusi soal yang dominan pada kategori kesukaran sedang sebesar 70% memperlihatkan kualitas perancangan instrumen yang seimbang dan sesuai dengan kebutuhan pembelajaran. Proporsi ini memungkinkan diferensiasi penilaian karena memfasilitasi siswa dengan tingkat kemampuan yang beragam. Penelitian *Anderson & Krathwohl* (2001) menegaskan bahwa kombinasi antara soal mudah, sedang, dan menantang mempermudah guru mengidentifikasi kesenjangan pemahaman siswa, sekaligus meningkatkan efektivitas asesmen berbasis HOTS.

Kualitas pengecoh merupakan faktor penentu efektivitas instrumen pilihan ganda. Hasil analisis menunjukkan bahwa 20 opsi (13,33%) memerlukan revisi agar lebih optimal berfungsi sebagai distraktor. Penelitian *Haladyna & Rodriguez* (2013) menegaskan bahwa pengecoh yang baik harus dibuat berdasarkan pola kesalahan umum yang sering dilakukan siswa. Oleh karena itu, perbaikan opsi jawaban yang kurang efektif dalam penelitian ini dapat meningkatkan daya ukur instrumen sekaligus memperkuat kesesuaian soal dengan indikator HOTS.

Penelitian ini memberikan kontribusi praktis bagi guru dan pengembang kurikulum dalam meningkatkan kualitas asesmen berbasis HOTS di tingkat sekolah menengah. Instrumen yang dihasilkan memenuhi standar evaluasi internasional karena memiliki validitas, reliabilitas, dan efektivitas pengecoh yang memadai. Hasil ini mendukung laporan *OECD (2021)* yang menegaskan pentingnya asesmen berbasis keterampilan analitis dan sintesis untuk mempersiapkan peserta didik menghadapi era globalisasi. Temuan penelitian ini dapat menjadi dasar pengembangan instrumen pembelajaran pada mata pelajaran geografi dan bidang lain secara lebih sistematis dan adaptif.

## SIMPULAN

Kesimpulan, hasil analisis terhadap 30 butir soal geografi yang diujikan kepada 38 responden menunjukkan bahwa 27 soal (90%) tergolong valid, sedangkan reliabilitas instrumen sangat tinggi dengan nilai 0,913. Analisis tingkat kesukaran mengungkapkan bahwa 30% soal berada pada kategori mudah, 70% kategori sedang, dan tidak ada soal yang tergolong sukar, di mana soal dengan tingkat kesukaran sedang dinilai paling ideal karena dapat mengukur kompetensi siswa secara lebih proporsional. Dari segi daya pembeda, 66% soal berada pada kategori sedang, 30% baik, dan 3% kuat, menunjukkan bahwa instrumen mampu membedakan siswa berdasarkan tingkat kemampuan mereka. Sementara itu, dari 150 opsi jawaban, terdapat 8 opsi (5,33%) yang ditolak dan 20 opsi (13,33%) yang direvisi karena dinilai kurang efektif sebagai pengecoh. Secara keseluruhan, temuan ini menegaskan bahwa instrumen yang dikembangkan telah memenuhi standar kualitas evaluasi dari aspek validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh, meskipun perbaikan pada beberapa opsi jawaban masih diperlukan agar instrumen dapat berfungsi secara optimal. Hasil penelitian ini diharapkan menjadi referensi dalam pengembangan dan analisis soal

berkualitas, sekaligus menjadi acuan bagi pendidik dalam merancang instrumen penilaian yang sesuai dengan kebutuhan pembelajaran abad ke-21, serta membuka peluang untuk penyempurnaan desain instrumen pada penelitian selanjutnya.

## DAFTAR RUJUKAN

- Adnan, A., Suwandi, S., Nurkamto, J., & Setiawan, B. (2019). Teacher competence in authentic and integrative assessment in Indonesian language learning. *International Journal of Instruction*, 12(1), 701–716. <https://doi.org/10.29333/iji.2019.12145a>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Anis, H. (n.d.). *Proses kognitif dalam taksonomi Bloom revisi Anderson dan Krathwohl*. Diakses pada 12 Juni 2023 dari <https://hermananis.com/proses-kognitif-dalam-taksonomi-bloom-revisi-anderson-dan-krathwohl/>
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Alexandria, VA: ASCD.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). London: Routledge. <https://doi.org/10.4324/9781315456539>
- Faizah, A., & Sutopo, D. (2021). The implementation of teachers' pedagogical and professional competence in authentic assessment. *English Education Journal*, 11(4), 539–549. <https://doi.org/10.15294/eej.v11i1.48823>
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items* (3rd ed.). New York: Routledge.
- Hidayati, A., Bentri, A., & Rahmi, U. (2017). Analyzing the issues in the implementation of authentic assessment in the 2013 curriculum. *Al-Ta'lim Journal*, 24(1), 53–59. <https://doi.org/10.15548/jt.v24i1.256>
- Indriyani, D., Artini, L., & Adnyani, L. (2021). The analysis of assessment instruments in English teacher lesson plans at junior high school in Singaraja Bali. *Journal of Research on English and Language Learning (J-Reall)*, 2(2), 83–93. <https://doi.org/10.33474/j-reall.v2i2.10118>
- Indriani, R. (2019). Authentic assessment awareness for assessing students' speaking ability in a suburban Indonesian EFL classroom. *Proceedings of the 11th Conference on Applied Linguistics (Conaplin 2018)*. <https://doi.org/10.2991/conaplin-18.2019.230>
- Jati, P., Ruhimat, M., & Logayah, D. S. (2023). Analisis kualitas butir soal geografi. *Jurnal Pendidikan Geosfer*, 8(2), 287–298.
- Marzano, R. J., & Heflebower, T. (2016). *Teaching and assessing 21st-century skills*. Bloomington, IN: Marzano Research Laboratory.
- Muluki, A. (2020). *Analisis kualitas butir tes semester ganjil mata pelajaran IPA kelas IV MI* [Skripsi tidak diterbitkan]. Universitas (tidak disebutkan).
- OECD. (2021). *PISA 2021 results: Learning, equity and assessment frameworks*. Paris: OECD Publishing. <https://doi.org/10.1787/pisa2021>

- 
- Ruhimat, M. (2019). Kompetensi pembuatan instrumen pengukuran hasil belajar oleh guru IPS SMP di Kota Bandung. *Lentera Pendidikan: Jurnal Ilmu Tarbiyah dan Keguruan*, 21(2), 176–187. <https://doi.org/10.24252/lp.2018v21n2i4>
- Rukmini, D., & Saputri, L. (2017). The authentic assessment to measure students' English productive skills based on 2013 curriculum. *Indonesian Journal of Applied Linguistics*, 7(2), 25–35. <https://doi.org/10.17509/ijal.v7i2.8128>
- Setiawati, N., Marsakawati, N., & Ramendra, D. (2024). Authentic assessment in action: English teachers' perception and practices for assessing students' productive skills at senior high school. *Indonesian Journal of Educational Development (IJED)*, 5(3), 366–375. <https://doi.org/10.59672/ijed.v5i3.4219>
- Sugandi, D., & Ruhimat, M. (2020). Assessment of geography for 21st century. *Jurnal Georaflesia: Artikel Ilmiah Pendidikan Geografi*, 5(1), 1–12.
- Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st-century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44(3), 299–321. <https://doi.org/10.1080/00220272.2012.668938>
- Wiliam, D. (2018). *Embedded formative assessment* (2nd ed.). Bloomington, IN: Solution Tree Press.
- Zaim, M., Mukhaiyar, M., & Syahriati, M. (2017). The need for authentic assessment for speaking skills at junior high school. *Proceedings of the 5th International Seminar on English Language and Teaching (ISELT 2017)*. <https://doi.org/10.2991/iselt-17.2017.1>