



MENJAMIN KETEPATAN INTERPRETASI HASIL BELAJAR: ANALISIS KOMPREHENSIF UJI VALIDITAS DALAM EVALUASI PEMBELAJARAN

Muhammad Rizqo Qolbiy^{1*}, & Dina Hermina²

^{*1&2} Universitas Islam Negeri Antasari Banjarmasin,

*e-mail: rizqoqolbiy@gmail.com

Submit Tgl: 05-Juni-2026

Diterima Tgl: 08-Juni-2026

Diterbitkan Tgl: 11-Juni-2026

Abstrak: Evaluasi pembelajaran merupakan komponen penting dalam sistem pendidikan yang berfungsi untuk mengukur ketercapaian tujuan pembelajaran. Keakuratan hasil evaluasi sangat ditentukan oleh kualitas instrumen yang digunakan, salah satunya melalui pemenuhan aspek validitas. Artikel ini bertujuan mengkaji secara komprehensif konsep, jenis, prosedur, dan implementasi uji validitas dalam evaluasi pembelajaran. Penelitian menggunakan metode studi kepustakaan (*library research*) dengan menelaah berbagai literatur ilmiah, buku, artikel jurnal, dan dokumen akademik yang berkaitan dengan validitas instrumen pendidikan. Hasil kajian menunjukkan bahwa validitas tidak hanya berkaitan dengan ketepatan butir instrumen, tetapi juga menyangkut ketepatan interpretasi dan penggunaan skor yang dihasilkan. Bentuk validitas yang umum digunakan meliputi validitas isi, validitas konstruk, dan validitas kriteria. Selain itu, perkembangan evaluasi pendidikan modern menempatkan validitas sebagai proses pengumpulan bukti yang berkelanjutan untuk mendukung penggunaan instrumen secara tepat. Oleh karena itu, pelaksanaan uji validitas menjadi langkah fundamental dalam menjamin kualitas evaluasi pembelajaran, meningkatkan akurasi pengambilan keputusan pendidikan, dan menghasilkan informasi yang dapat dipertanggungjawabkan secara ilmiah.

Kata Kunci: Validitas, Evaluasi Pembelajaran, Instrumen Asesmen, Pengukuran Pendidikan, Kualitas Instrumen

Abstract: Learning evaluation is an essential component of the educational system that aims to measure the achievement of learning objectives. The accuracy of evaluation results largely depends on the quality of the instruments employed, particularly their validity. This article aims to comprehensively examine the concepts, types, procedures, and implementation of validity testing in learning evaluation. The study employed a library research approach by reviewing scientific literature, academic books, journal articles, and educational assessment documents related to instrument validity. The findings indicate that validity is not merely associated with the accuracy of test items but also concerns the appropriateness of score interpretation and utilization. Common forms of validity include content validity, construct validity, and criterion-related validity. Furthermore, contemporary educational assessment perspectives regard validity as an ongoing process of collecting evidence to support the intended use of assessment results. Therefore, validity testing constitutes a fundamental step in ensuring the quality of learning evaluation, improving the accuracy of educational decision-making, and generating scientifically accountable information.

Keywords: Strategy, Instructional Design, Islamic Religious Education, Learning Planning, Educational Innovation

Cara mengutip Qolbiy, M. R., & Hermina, D. (2026). Menjamin Ketepatan Interpretasi Hasil Belajar: Analisis Komprehensif Uji Validitas dalam Evaluasi Pembelajaran. *JIS: Journal Islamic Studies*, 4(2), 374–381. <https://doi.org/10.71456/jis.v4i2.1972>

PENDAHULUAN

Evaluasi pembelajaran merupakan salah satu komponen esensial dalam sistem pendidikan yang berfungsi untuk memperoleh informasi mengenai tingkat pencapaian tujuan pembelajaran. Melalui kegiatan evaluasi, pendidik dapat mengetahui perkembangan peserta didik, efektivitas proses pembelajaran, serta tingkat keberhasilan program pendidikan yang dilaksanakan (Toha et al., 2026). Dalam konteks pendidikan modern, evaluasi tidak lagi dipahami sekadar sebagai kegiatan pemberian nilai, melainkan sebagai proses sistematis yang menghasilkan data untuk mendukung berbagai keputusan pendidikan (Sudarman et al., 2025). Oleh karena itu, kualitas hasil evaluasi sangat bergantung pada kualitas instrumen yang digunakan dalam proses pengukuran.

Meskipun evaluasi pembelajaran memiliki peran yang sangat strategis, berbagai penelitian menunjukkan bahwa masih banyak instrumen evaluasi yang digunakan tanpa melalui proses pengujian kualitas yang memadai (Saputra, 2025). Dalam praktiknya, guru sering kali menyusun soal atau instrumen asesmen berdasarkan pengalaman dan intuisi tanpa melakukan analisis empiris terhadap kualitas instrumen tersebut. Akibatnya, hasil evaluasi yang diperoleh berpotensi mengandung kesalahan pengukuran yang dapat memengaruhi ketepatan interpretasi terhadap kemampuan peserta didik. Kondisi ini menunjukkan adanya kesenjangan antara tuntutan teoretis pengembangan instrumen yang berkualitas dengan praktik evaluasi yang terjadi di lapangan.

Salah satu aspek fundamental yang menentukan kualitas instrumen evaluasi adalah validitas. Konsep validitas menempati posisi sentral dalam teori pengukuran pendidikan karena berkaitan dengan sejauh mana instrumen mampu mengukur apa yang seharusnya diukur. Dalam perkembangan ilmu evaluasi pendidikan, validitas tidak lagi dipahami hanya sebagai karakteristik instrumen, melainkan sebagai derajat ketepatan interpretasi terhadap skor yang dihasilkan (Yusup, 2018). Dengan demikian, validitas menjadi dasar utama dalam memastikan bahwa keputusan yang diambil berdasarkan hasil evaluasi memiliki landasan ilmiah yang kuat.

Perkembangan paradigma asesmen abad ke-21 semakin menuntut penggunaan instrumen yang mampu mengukur kompetensi peserta didik secara autentik dan komprehensif. Penilaian tidak hanya berfokus pada aspek kognitif, tetapi juga mencakup keterampilan berpikir kritis, kreativitas, kolaborasi, komunikasi, dan berbagai kompetensi lainnya (Kalyana et al., 2025). Kompleksitas tersebut menyebabkan proses validasi instrumen menjadi semakin penting karena instrumen harus mampu merepresentasikan konstruk yang diukur secara akurat. Kesalahan dalam proses validasi dapat mengakibatkan kesimpulan yang tidak tepat mengenai kemampuan peserta didik maupun efektivitas program pembelajaran.

Berdasarkan uraian tersebut, kajian mengenai uji validitas dalam evaluasi pembelajaran menjadi sangat relevan untuk dilakukan. Selain memberikan pemahaman konseptual mengenai hakikat validitas, kajian ini juga penting untuk menjelaskan berbagai jenis validitas, prosedur pengujiannya, serta implikasinya terhadap kualitas evaluasi pembelajaran. Dengan memahami konsep validitas secara komprehensif, pendidik dan peneliti pendidikan diharapkan mampu mengembangkan instrumen evaluasi yang lebih berkualitas sehingga hasil pengukuran yang diperoleh dapat digunakan sebagai dasar pengambilan keputusan pendidikan yang tepat, objektif, dan dapat dipertanggungjawabkan.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan metode studi kepustakaan (*library research*). Metode ini dipilih karena tujuan penelitian adalah menganalisis dan mensintesis berbagai konsep, teori, dan temuan penelitian yang berkaitan dengan uji validitas dalam evaluasi pembelajaran (Rocco & Plakhotnik, 2009). Data penelitian diperoleh dari berbagai sumber literatur yang relevan, seperti buku ilmiah, artikel jurnal, prosiding, serta dokumen akademik yang membahas validitas instrumen pendidikan (Sidiq, 2019).

Pengumpulan data dilakukan melalui teknik dokumentasi dengan menelusuri berbagai publikasi ilmiah yang diterbitkan dalam rentang waktu terbaru maupun literatur klasik yang masih relevan (Harahap, 2020). Literatur yang dipilih memenuhi kriteria kredibilitas akademik, relevansi dengan topik penelitian, serta memiliki kontribusi terhadap pengembangan konsep validitas dalam evaluasi pembelajaran.

Analisis data dilakukan melalui tiga tahapan utama, yaitu reduksi data, penyajian data, dan penarikan kesimpulan (Sidiq, 2019). Pada tahap reduksi, berbagai informasi yang diperoleh dari sumber literatur diseleksi berdasarkan relevansinya dengan fokus kajian. Selanjutnya, data disajikan secara sistematis sesuai dengan tema-tema pembahasan yang telah ditentukan. Tahap akhir dilakukan dengan menginterpretasikan berbagai temuan untuk menghasilkan pemahaman yang komprehensif mengenai konsep, jenis, prosedur, dan implikasi uji validitas dalam evaluasi pembelajaran.

HASIL DAN PEMBAHASAN

Hakikat dan Kedudukan Validitas dalam Evaluasi Pembelajaran

Validitas merupakan salah satu konsep paling fundamental dalam bidang evaluasi dan pengukuran pendidikan. Secara umum, validitas merujuk pada tingkat ketepatan suatu instrumen dalam mengukur apa yang seharusnya diukur (Saputri & Larasati, 2023). Dalam konteks evaluasi pembelajaran, validitas menjadi dasar untuk menentukan apakah skor yang diperoleh peserta didik benar-benar mencerminkan kompetensi, pengetahuan, keterampilan, atau sikap yang hendak diukur. Instrumen yang tidak valid dapat menghasilkan informasi yang menyesatkan sehingga keputusan pendidikan yang diambil berdasarkan hasil pengukuran tersebut menjadi tidak tepat (Fernando et al., 2023). Oleh karena itu, validitas sering disebut sebagai “jiwa” dari suatu instrumen evaluasi. Tanpa validitas yang memadai, instrumen kehilangan fungsinya sebagai alat ukur yang dapat dipercaya. Berbagai kajian dalam bidang evaluasi pendidikan menempatkan validitas sebagai syarat utama yang harus dipenuhi sebelum aspek-aspek lain seperti reliabilitas, daya pembeda, maupun tingkat kesukaran dianalisis lebih lanjut.

Dalam perkembangan teori pengukuran modern, konsep validitas mengalami perubahan yang cukup signifikan. Pada masa awal perkembangan evaluasi pendidikan, validitas dipahami sebagai karakteristik yang melekat pada instrumen. Namun, para ahli pengukuran kontemporer memandang bahwa validitas sesungguhnya berkaitan dengan ketepatan interpretasi dan penggunaan skor yang dihasilkan oleh instrumen tersebut (Hermawan et al., 2025). Dengan kata lain, suatu instrumen tidak dapat secara mutlak disebut valid atau tidak valid, melainkan bukti-bukti validitas harus terus dikumpulkan untuk mendukung penggunaan instrumen dalam konteks tertentu. Pandangan ini menunjukkan bahwa validitas merupakan proses yang berkelanjutan dan memerlukan evaluasi secara terus-menerus sesuai dengan tujuan penggunaannya.

Dalam praktik pendidikan, validitas memiliki posisi strategis karena menjadi dasar dalam berbagai pengambilan keputusan. Hasil evaluasi digunakan untuk menentukan

kenaikan kelas, kelulusan, penempatan peserta didik, diagnosis kesulitan belajar, hingga evaluasi efektivitas kurikulum (Siregar, 2025). Jika instrumen yang digunakan tidak memiliki validitas yang memadai, maka seluruh keputusan tersebut berpotensi mengandung kesalahan. Misalnya, peserta didik yang sebenarnya menguasai materi dapat dinilai tidak kompeten karena instrumen gagal mengukur kompetensi yang sesungguhnya. Sebaliknya, peserta didik yang belum menguasai materi dapat memperoleh skor tinggi akibat kelemahan instrumen yang digunakan.

Urgensi validitas semakin meningkat pada era pendidikan abad ke-21. Saat ini, pembelajaran tidak hanya menekankan penguasaan materi, tetapi juga kemampuan berpikir kritis, kreativitas, komunikasi, kolaborasi, literasi digital, dan pemecahan masalah (Sari et al., 2020). Kompetensi-kompetensi tersebut memiliki konstruk yang kompleks sehingga memerlukan instrumen yang dirancang secara cermat dan divalidasi secara sistematis. Oleh sebab itu, pemahaman yang mendalam mengenai validitas menjadi kebutuhan penting bagi guru, dosen, peneliti pendidikan, maupun pengembang instrumen asesmen agar hasil evaluasi benar-benar mencerminkan kemampuan peserta didik yang sebenarnya.

Jenis-Jenis Validitas dalam Pengembangan Instrumen Evaluasi

Validitas dalam evaluasi pembelajaran dapat diklasifikasikan ke dalam beberapa jenis berdasarkan sumber bukti yang digunakan untuk mendukung ketepatan interpretasi hasil pengukuran. Kurniawan et al. (2022) menyatakan bahwa secara umum, jenis validitas yang paling banyak digunakan dalam pengembangan instrumen pendidikan meliputi validitas isi (*content validity*), validitas konstruk (*construct validity*), dan validitas kriteria (*criterion-related validity*). Ketiga jenis validitas tersebut saling melengkapi dan memberikan gambaran yang lebih komprehensif mengenai kualitas suatu instrumen evaluasi \.

Validitas isi merupakan bentuk validitas yang menunjukkan sejauh mana butir-butir instrumen telah mewakili seluruh domain materi atau kompetensi yang hendak diukur. Dalam evaluasi pembelajaran, validitas isi biasanya diperoleh melalui penilaian para ahli (*expert judgment*). Para ahli akan menelaah kesesuaian antara indikator pembelajaran, kisi-kisi instrumen, dan butir soal yang disusun. Semakin tinggi tingkat kesesuaian antara instrumen dan kompetensi yang diukur, semakin tinggi pula validitas isi yang dimiliki instrumen tersebut. Oleh karena itu, penyusunan kisi-kisi yang baik menjadi langkah awal yang sangat penting dalam proses validasi instrument (Hendryadi, 2017).

Jenis validitas berikutnya adalah validitas konstruk. Validitas konstruk berkaitan dengan sejauh mana instrumen benar-benar mengukur konstruk teoretis yang menjadi dasar penyusunannya. Konstruk dapat berupa kemampuan berpikir kritis, motivasi belajar, kreativitas, kecerdasan emosional, atau kompetensi lainnya yang bersifat abstrak. Aziz (2022) menjelaskan pengujian validitas konstruk umumnya dilakukan melalui analisis statistik seperti *Exploratory Factor Analysis* (EFA) dan *Confirmatory Factor Analysis* (CFA). Dalam penelitian pendidikan modern, validitas konstruk sering dianggap sebagai bentuk validitas yang paling kuat karena memberikan bukti empiris mengenai struktur teoritis instrumen. Faktor loading yang memadai dan kesesuaian model yang baik menjadi indikator utama dalam menilai validitas konstruk.

Adapun validitas kriteria mengacu pada hubungan antara skor instrumen dengan ukuran eksternal yang relevan. Validitas ini dibedakan menjadi validitas konkuren (*concurrent validity*) dan validitas prediktif (*predictive validity*). Validitas konkuren

digunakan ketika skor instrumen dibandingkan dengan ukuran lain yang diperoleh pada waktu yang sama, sedangkan validitas prediktif digunakan untuk memprediksi kinerja atau hasil yang akan datang (Buka et al., 2025). Sebagai contoh, skor tes masuk perguruan tinggi dapat diuji validitas prediktifnya dengan membandingkannya terhadap prestasi akademik mahasiswa pada semester berikutnya. Semakin kuat hubungan antara skor instrumen dan kriteria eksternal, semakin tinggi validitas kriteria yang dimiliki instrumen tersebut.

Prosedur dan Teknik Uji Validitas Instrumen Pembelajaran

Pelaksanaan uji validitas instrumen pembelajaran dilakukan melalui serangkaian tahapan yang sistematis. Tahap pertama adalah penyusunan spesifikasi instrumen berdasarkan tujuan pembelajaran dan indikator kompetensi yang telah ditetapkan (Shalihah et al., 2026). Pada tahap ini, pengembang instrumen menyusun kisi-kisi yang berfungsi sebagai pedoman dalam merumuskan butir-butir soal atau pernyataan. Kisi-kisi yang baik akan memastikan bahwa setiap indikator kompetensi memperoleh representasi yang proporsional dalam instrumen yang dikembangkan.

Tahap kedua adalah validasi ahli atau *expert judgment*. Dalam tahap ini, instrumen yang telah disusun ditelaah oleh para pakar sesuai bidang keahlian masing-masing. Para validator memberikan penilaian mengenai relevansi isi, kejelasan bahasa, kesesuaian indikator, dan ketepatan format instrumen (Permatasari & Purwandari, 2025). Hasil penilaian ahli biasanya dianalisis menggunakan indeks validitas seperti Aiken's V atau *Content Validity Index* (CVI). Apabila ditemukan kelemahan pada instrumen, maka dilakukan revisi sebelum instrumen diuji coba kepada responden (Kania et al., 2024).

Tahap ketiga adalah uji coba lapangan (*pilot testing*). Instrumen yang telah direvisi kemudian diberikan kepada sejumlah responden yang memiliki karakteristik serupa dengan populasi penelitian. Data hasil uji coba digunakan untuk menganalisis validitas empiris setiap butir instrumen. Pada instrumen tes, validitas butir sering dihitung menggunakan korelasi *Product Moment Pearson* antara skor butir dengan skor total (Yusuf, 2017). Secara umum, suatu butir dinyatakan valid apabila memiliki koefisien korelasi yang lebih besar daripada nilai kritis yang ditentukan. Semakin tinggi korelasi antara skor butir dan skor total, semakin baik kemampuan butir tersebut dalam mengukur konstruk yang dimaksud.

Pada instrumen yang lebih kompleks, terutama instrumen non-tes seperti angket sikap, motivasi, atau persepsi, analisis validitas dapat dilakukan menggunakan pendekatan faktor. Kamila dan Widodo (2025) menjelaskan analisis faktor eksploratori digunakan untuk mengidentifikasi struktur dimensi yang muncul dari data, sedangkan analisis faktor konfirmatori digunakan untuk menguji kesesuaian model teoritis dengan data empiris. Berbagai indeks seperti RMSEA, CFI, TLI, (Xia & Yang, 2019) dan Chi-Square digunakan untuk menilai kualitas model (Pandis, 2016). Penggunaan teknik-teknik statistik modern tersebut menunjukkan bahwa proses validasi instrumen saat ini semakin menekankan pentingnya bukti empiris yang kuat dalam mendukung penggunaan hasil pengukuran.

Implikasi Validitas terhadap Kualitas Hasil Evaluasi dan Pengambilan Keputusan Pendidikan

Validitas memiliki implikasi yang sangat besar terhadap kualitas hasil evaluasi pembelajaran. Instrumen yang valid akan menghasilkan data yang lebih akurat sehingga mampu menggambarkan kondisi peserta didik secara objektif. Sebaliknya, instrumen

yang tidak valid berpotensi menghasilkan informasi yang bias dan menyesatkan. Kesalahan dalam pengukuran dapat berdampak pada kesalahan interpretasi hasil belajar yang pada akhirnya memengaruhi berbagai keputusan pendidikan yang diambil oleh guru maupun lembaga pendidikan.

Dalam konteks pembelajaran, validitas berkontribusi terhadap terciptanya asesmen yang adil (*fair assessment*) (Krans et al., 2022). Ketika instrumen benar-benar mengukur kompetensi yang telah diajarkan, peserta didik memperoleh kesempatan yang sama untuk menunjukkan kemampuan mereka. Hal ini penting untuk menjaga prinsip keadilan dan akuntabilitas dalam proses pendidikan. Instrumen yang valid juga membantu guru mengidentifikasi kekuatan dan kelemahan peserta didik secara lebih tepat sehingga program remedial maupun pengayaan dapat dirancang secara efektif.

Selain berdampak pada peserta didik, validitas juga berpengaruh terhadap kualitas evaluasi program pendidikan secara keseluruhan. Hasil evaluasi sering digunakan sebagai dasar untuk menilai keberhasilan kurikulum, efektivitas metode pembelajaran, serta kinerja lembaga pendidikan. Apabila data yang digunakan berasal dari instrumen yang tidak valid, maka kesimpulan yang dihasilkan pun berpotensi keliru. Oleh sebab itu, validitas tidak hanya berkaitan dengan kualitas instrumen, tetapi juga berkaitan dengan kualitas kebijakan pendidikan yang disusun berdasarkan hasil evaluasi tersebut.

Dalam perspektif yang lebih luas, Wingfield et al. (2004) menjelaskan bahwa validitas merupakan bagian dari upaya mewujudkan budaya evaluasi yang ilmiah dan berbasis bukti (*evidence-based evaluation*). Semakin tinggi kualitas validitas instrumen yang digunakan, semakin tinggi pula tingkat kepercayaan terhadap hasil evaluasi yang diperoleh. Dengan demikian, validitas bukan sekadar prosedur teknis dalam pengembangan instrumen, melainkan fondasi utama yang menentukan kredibilitas seluruh proses evaluasi pembelajaran dan pengambilan keputusan pendidikan.

KESIMPULAN

Validitas merupakan aspek paling fundamental dalam evaluasi pembelajaran karena menentukan ketepatan instrumen dalam mengukur kompetensi yang menjadi sasaran pengukuran. Perkembangan teori evaluasi modern menunjukkan bahwa validitas tidak hanya berkaitan dengan kualitas instrumen, tetapi juga dengan ketepatan interpretasi dan penggunaan skor yang dihasilkan. Bentuk-bentuk validitas yang umum digunakan meliputi validitas isi, validitas konstruk, dan validitas kriteria, yang masing-masing memberikan bukti berbeda mengenai kualitas instrumen.

Proses uji validitas dilakukan melalui tahapan yang sistematis, mulai dari penyusunan kisi-kisi, validasi ahli, uji coba lapangan, hingga analisis statistik empiris. Pelaksanaan prosedur tersebut bertujuan untuk memastikan bahwa instrumen benar-benar mampu menghasilkan data yang akurat dan dapat dipertanggungjawabkan. Hasil kajian menunjukkan bahwa validitas memiliki implikasi yang sangat luas terhadap kualitas evaluasi pembelajaran, keadilan asesmen, efektivitas pembelajaran, serta ketepatan pengambilan keputusan pendidikan.

Oleh karena itu, setiap instrumen evaluasi yang digunakan dalam pembelajaran harus melalui proses validasi yang memadai sebelum diterapkan. Semakin kuat bukti validitas yang dimiliki suatu instrumen, semakin tinggi pula tingkat kepercayaan terhadap hasil pengukuran yang dihasilkan. Dengan demikian, validitas menjadi fondasi utama dalam mewujudkan evaluasi pembelajaran yang objektif, akurat, dan berbasis bukti ilmiah.

DAFTAR PUSTAKA

- Aziz, H. (2022). *Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to measure the validity and reliability constructs of historical thinking skills, TPACK. International Journal of Education, Psychology and Counseling.*
- Buka, N., Jayamin, M. A., Kholis, N., et al. (2025). Peran validitas dan rehabilitas dalam evaluasi pembelajaran. *Socius: Jurnal Penelitian Ilmu Sosial dan Pendidikan.*
- Fernando, D. A., Hartatiana, H., et al. (2023). Pentingnya validitas dan reliabilitas instrumen evaluasi pada pembelajaran pendidikan agama Islam. *Raudhah Proud To Be Professionals: Jurnal Tarbiyah Islamiyah.*
- Harahap, N. (2020). *Penelitian kualitatif* (H. Sazali, Ed.). Medan.
- Hendryadi. (2017). Validitas isi: Tahap awal pengembangan kuesioner. *Jurnal Riset Manajemen dan Bisnis*, 2(2), 169–178.
- Hermawan, M. H., Soeharno, S., Supriadi, W., et al. (2025). Pengembangan instrumen pengukuran keterampilan kolaborasi terintegrasi nilai pendidikan Islam abad 21 siswa sekolah menengah pertama Islam. *Journal of Education Research.*
- Kalyana, D., Hayatunnufus, G., Fauziah, F. H., et al. (2025). Implementasi asesmen autentik berbasis kompetensi dalam pembelajaran abad ke-21: Penelitian. *Jurnal Edukasi, Riset, dan Inovasi Masyarakat.*
- Kamila, A., & Widodo, P. B. (2025). Validitas konstruk dalam psikometri: Peran analisis faktor eksploratori dan konfirmatori. *Jurnal Psikologi Pendidikan dan Konseling.*
- Kania, N., Kusumah, Y. S., Dahlan, J. A., et al. (2024). Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem solving. *Research and Evaluation in Education*, 10(1).
- Krans, N. A., Ammar, A., Nymark, P., Willighagen, E. L., et al. (2022). FAIR assessment tools: Evaluating use and performance. *NanoImpact*, 25, 100374.
- Kurniawan, A., Febriant, A. N., Hardianti, T., Ichsan, I., et al. (2022). *Evaluasi pembelajaran.* Universitas Kristen Indonesia.
- Pandis, N. (2016). The chi-square test. *American Journal of Orthodontics and Dentofacial Orthopedics*, 150(5), 898–899.
- Permatasari, A. A., & Purwandari, E. (2025). Validitas isi alat ukur dukungan sosial melalui expert judgement pada mahasiswa terapi wicara. *Jurnal Terapi Wicara dan Bahasa.*
- Rocco, T. S., & Plakhotnik, M. S. (2009). Literature reviews, conceptual frameworks, and theoretical frameworks: Terms, functions, and distinctions. *Human Resource Development Review*, 8(1), 120–130.
- Salihah, I., Raswan, R., & Ridlo, U. (2026). Uji validitas dan reliabilitas. *Luxfia: Journal of Islamic Education.*
- Saputra, A. (2025). Pengembangan instrumen evaluasi. *Ar-Raudah: Jurnal Pendidikan dan Keagamaan.*

- Saputri, H. A. S., & Larasati, N. J. (2023). Analisis instrumen asesmen: Validitas, reliabilitas, tingkat kesukaran dan daya beda butir soal. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*.
- Sari, L. P. N., Fajarianingtyas, D. A., et al. (2020). Validitas instrumen penilaian kemampuan berpikir kritis melalui model problem based learning menuju pembelajaran IPA abad ke-21. *LENSA (Lentera Sains): Jurnal Pendidikan IPA*.
- Sidiq, U. (2019). *Metode penelitian kualitatif di bidang pendidikan* (A. Mujahidin, Ed.). CV. Nata Karya.
- Siregar, T. (2025). Mengapa validitas dan reliabilitas penting dalam L&D (pembelajaran dan pengembangan): Apa yang dapat Anda lakukan tentang hal ini. *Jurnal Pemikiran dan Pengembangan Pendidikan*.
- Sudarman, S., Listiani, H., Judijanto, L., Warikar, A., et al. (2025). *Evaluasi pendidikan abad 21: Inovasi, spiritualitas, dan relevansi*. CV. Media Sains Indonesia.
- Toha, M., Munirah, M., & Mulkiyah, S. (2026). Evaluasi pendidikan: Antara konsep, urgensi dan efektivitasnya dalam meningkatkan mutu pendidikan. *Jurnal Multidisiplin*.
- Wardati. (2025). Asesmen otentik dalam konteks pendidikan abad 21. Dalam *Sistem asesmen dan evaluasi pendidikan: Paradigma, model, implementasi, dan tantangan di era abad 21*.
- Wingfield, K., Matheson, G. O., et al. (2004). Preparticipation evaluation: An evidence-based review. *Clinical Journal of Sport Medicine*, 14(3), 109–122.
- Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, 51(1), 409–428.
- Yusuf, A. M. (2017). *Asesmen dan evaluasi pendidikan*. Kencana
- Yusup, F. (2018). Uji validitas dan reliabilitas instrumen penelitian kuantitatif. *Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(1), 17–23.