

VALIDITAS DAN RELIABILITAS INSTRUMEN EVALUASI

Husain Alfath, Wawan Arbeni*, Ahmad Alma Ariz, Ayu Wardani, Karisma Annisa, M. Hendryk Nasution, Elia Fitri, Emi Pratiwi

wawanarbeni@insan.ac.id

Institut Syekh Abdul Halim Hasan Binjai

ARTICLE INFO

Keywords: educational evaluation, standardized tests, non-standard tests, learning assessment

Received : ...

Revised : ...

Accepted : ...

©2023 The Author(s): This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by-sa/4.0/)



ABSTRACT

Validity and reliability are two fundamental aspects that determine the quality of evaluation instruments in the education and writing process. Evaluation instruments that do not meet these two criteria have the potential to produce biased and inaccurate data, which can affect the accuracy of learning decisions. This paper aims to comprehensively examine the concepts of validity and reliability of evaluation instruments, including the types of validity and reliability, factors that influence them, testing methods, and techniques for processing evaluation scores. This article uses a library research approach by analyzing various literature sources in the form of textbooks, national and international scientific journals, and relevant academic documents. The data were analyzed descriptively and qualitatively through the process of organizing, grouping, and synthesizing the main concepts. The results of the study show that the quality of evaluation is not only determined by the validity and reliability of the instruments, but also by the accuracy in processing and interpreting scores, such as converting raw scores into letter grades and Z-scores. Therefore, validity, reliability, and score transformation are an interrelated unit in producing learning evaluations that are objective, fair, and scientifically accountable.

INTRODUCTION

Evaluasi merupakan komponen yang tidak terpisahkan dari proses pendidikan, karena melalui evaluasi pendidik dapat mengetahui sejauh mana tujuan pembelajaran telah tercapai. Evaluasi tidak hanya berfungsi sebagai alat untuk mengukur hasil belajar peserta didik, tetapi juga sebagai dasar pengambilan keputusan dalam perencanaan pembelajaran selanjutnya. Dalam konteks ini, kualitas evaluasi sangat bergantung pada instrumen yang digunakan. Instrumen evaluasi yang disusun secara asal tanpa memperhatikan kaidah pengukuran akan menghasilkan data yang bias, tidak akurat, dan berpotensi menyesatkan dalam menilai kemampuan peserta didik.

Instrumen evaluasi pada dasarnya dirancang untuk mengukur kompetensi tertentu yang telah ditetapkan dalam tujuan pembelajaran. Kompetensi tersebut dapat berupa pengetahuan, keterampilan, maupun sikap. Agar instrumen benar-benar mampu mengukur apa yang seharusnya diukur, maka instrumen tersebut harus memenuhi persyaratan utama dalam pengukuran pendidikan, yaitu validitas dan reliabilitas. Tanpa adanya validitas dan reliabilitas yang memadai, hasil evaluasi tidak dapat dijadikan dasar yang kuat dalam menentukan keberhasilan pembelajaran maupun pencapaian kompetensi peserta didik.

Validitas berkaitan dengan sejauh mana suatu instrumen mampu mengukur apa yang seharusnya diukur sesuai dengan tujuan pengukuran. Instrumen yang valid akan memberikan gambaran yang tepat mengenai kemampuan atau karakteristik peserta didik yang dinilai. Sebaliknya, instrumen yang tidak valid dapat mengukur aspek lain di luar tujuan yang telah ditetapkan, sehingga hasilnya menjadi tidak relevan. Oleh karena itu, validitas merupakan aspek fundamental yang harus diperhatikan sejak tahap perencanaan dan penyusunan instrumen evaluasi.

Selain validitas, reliabilitas juga menjadi syarat utama dalam menentukan kualitas suatu instrumen evaluasi. Reliabilitas berhubungan dengan konsistensi hasil pengukuran yang diperoleh dari suatu instrumen. Instrumen yang reliabel akan memberikan hasil yang relatif sama apabila digunakan dalam kondisi yang setara, meskipun pengukuran dilakukan pada waktu yang berbeda atau oleh penilai yang berbeda. Konsistensi ini penting agar hasil evaluasi dapat dipercaya dan tidak dipengaruhi oleh faktor kebetulan semata.

Dalam praktik pendidikan, masih banyak instrumen evaluasi yang digunakan tanpa melalui proses pengujian validitas dan reliabilitas secara sistematis. Hal

ini sering terjadi karena keterbatasan pemahaman pendidik terhadap konsep pengukuran pendidikan atau karena tuntutan administrasi yang harus dipenuhi dalam waktu singkat. Akibatnya, instrumen yang digunakan belum tentu mencerminkan kemampuan peserta didik secara akurat, sehingga keputusan yang diambil berdasarkan hasil evaluasi tersebut menjadi kurang tepat.

Instrumen evaluasi yang baik seharusnya tidak hanya mampu mengukur hasil belajar secara objektif, tetapi juga memberikan informasi yang bermakna bagi pendidik dan peserta didik. Informasi tersebut dapat digunakan untuk memperbaiki strategi pembelajaran, memberikan umpan balik kepada peserta didik, serta meningkatkan kualitas proses pembelajaran secara keseluruhan. Oleh karena itu, pemahaman yang mendalam mengenai validitas dan reliabilitas instrumen evaluasi menjadi kebutuhan yang sangat penting bagi pendidik, peneliti, maupun praktisi pendidikan.

Dalam konteks penulisan pendidikan, validitas dan reliabilitas instrumen memiliki peran yang sangat krusial. Data yang dikumpulkan melalui instrumen yang tidak valid dan tidak reliabel akan menghasilkan temuan penulisan yang lemah dan sulit dipertanggungjawabkan secara ilmiah. Oleh sebab itu, setiap penulisan yang menggunakan instrumen evaluasi wajib menyertakan proses pengujian validitas dan reliabilitas sebagai bagian dari metodologi penulisan. Hal ini bertujuan untuk menjamin bahwa data yang diperoleh benar-benar mencerminkan variabel yang diteliti.

Pengujian validitas dan reliabilitas tidak hanya terbatas pada tes tertulis, tetapi juga mencakup instrumen non-tes seperti angket, lembar observasi, dan rubrik penilaian. Setiap jenis instrumen memiliki karakteristik yang berbeda, sehingga pendekatan pengujian validitas dan reliabilitasnya pun dapat bervariasi. Oleh karena itu, pemahaman mengenai jenis-jenis validitas dan reliabilitas menjadi sangat penting agar pendidik dan peneliti dapat memilih metode pengujian yang sesuai dengan instrumen yang digunakan.

Selain pengujian instrumen, pengolahan hasil evaluasi juga memegang peranan penting dalam sistem penilaian. Skor mentah yang diperoleh peserta didik dari suatu tes belum dapat langsung memberikan makna yang komprehensif apabila tidak diolah dengan tepat. Pengolahan skor mentah menjadi nilai huruf, misalnya, merupakan salah satu cara untuk mengklasifikasikan tingkat pencapaian peserta didik berdasarkan kriteria tertentu. Proses ini membutuhkan pemahaman yang baik agar penilaian yang dihasilkan bersifat adil dan objektif.

Selain konversi ke dalam nilai huruf, skor mentah juga dapat diolah menjadi skor standar, seperti skor Z. Skor standar memungkinkan pendidik dan peneliti untuk membandingkan hasil belajar peserta didik dalam satu kelompok maupun antar kelompok yang berbeda. Dengan menggunakan skor Z, posisi relatif seorang peserta didik dalam distribusi nilai dapat diketahui secara lebih jelas, sehingga interpretasi hasil evaluasi menjadi lebih akurat dan informatif.

Penggunaan skor standar sangat bermanfaat dalam konteks evaluasi pendidikan, terutama ketika penilaian dilakukan dalam skala besar atau melibatkan banyak peserta didik. Skor standar membantu mengurangi bias penilaian yang mungkin muncul akibat perbedaan tingkat kesulitan tes atau variasi kemampuan antar kelompok. Oleh karena itu, pemahaman mengenai cara mengolah skor mentah menjadi skor standar menjadi bagian penting dalam sistem evaluasi yang berkualitas.

Namun demikian, pengolahan skor dan penggunaan skor standar akan kehilangan maknanya apabila instrumen yang digunakan tidak valid dan reliabel. Oleh karena itu, pengujian dan peningkatan validitas serta reliabilitas instrumen harus dilakukan secara berkelanjutan. Proses ini mencakup analisis butir soal, revisi instrumen, serta evaluasi terhadap hasil pengukuran yang telah dilakukan sebelumnya. Dengan demikian, kualitas instrumen evaluasi dapat terus ditingkatkan seiring dengan perkembangan kebutuhan pembelajaran.

Upaya meningkatkan validitas dan reliabilitas instrumen evaluasi juga menuntut keterampilan teknis dan pemahaman konseptual yang memadai. Pendidik dan peneliti perlu memahami prinsip-prinsip dasar statistik pendidikan serta teknik analisis yang relevan. Meskipun hal ini sering dianggap kompleks, pemahaman yang baik akan membantu dalam menghasilkan instrumen evaluasi yang lebih akurat dan dapat dipercaya.

Seiring dengan perkembangan kurikulum dan tuntutan pembelajaran abad ke-21, evaluasi pendidikan tidak lagi hanya berfokus pada aspek kognitif, tetapi juga mencakup keterampilan berpikir kritis, kreativitas, kolaborasi, dan komunikasi. Hal ini menuntut instrumen evaluasi yang lebih kompleks dan beragam. Oleh karena itu, validitas dan reliabilitas instrumen menjadi semakin penting untuk memastikan bahwa aspek-aspek tersebut dapat diukur secara tepat.

Dalam konteks penilaian berbasis kompetensi, instrumen evaluasi harus benar-benar selaras dengan capaian pembelajaran yang ditetapkan. Ketidaksesuaian antara tujuan pembelajaran, instrumen evaluasi, dan teknik penilaian akan menyebabkan hasil evaluasi tidak mencerminkan kompetensi yang sebenarnya.

Oleh karena itu, validitas isi dan validitas konstruk menjadi aspek yang sangat penting dalam penyusunan instrumen evaluasi berbasis kompetensi.

Dengan demikian, artikel ini bertujuan untuk membahas secara komprehensif konsep validitas dan reliabilitas instrumen evaluasi, mulai dari landasan teoretis hingga penerapannya dalam praktik penilaian pendidikan. Pembahasan ini diharapkan dapat memberikan kontribusi bagi pendidik, mahasiswa, dan peneliti dalam meningkatkan kualitas evaluasi pembelajaran serta menghasilkan penilaian yang lebih objektif, adil, dan dapat dipertanggungjawabkan secara ilmiah.

LITERATURE REVIEW

Instrumen evaluasi memegang peranan penting dalam proses penilaian hasil belajar karena kualitas data yang dihasilkan sangat bergantung pada ketepatan dan konsistensi alat ukur yang digunakan. Dalam konteks pendidikan, instrumen evaluasi dirancang untuk memperoleh informasi mengenai capaian belajar peserta didik secara objektif dan sistematis, baik pada ranah kognitif, afektif, maupun psikomotorik. Oleh sebab itu, pembahasan mengenai validitas dan reliabilitas menjadi aspek fundamental dalam kajian evaluasi pendidikan. Validitas instrumen evaluasi merujuk pada tingkat ketepatan suatu alat ukur dalam mengukur kemampuan, pengetahuan, sikap, atau keterampilan yang menjadi sasaran evaluasi. Instrumen yang valid mampu merepresentasikan secara akurat konstruk atau kompetensi yang hendak diukur sesuai dengan tujuan pembelajaran. Apabila instrumen tidak selaras dengan indikator pembelajaran, maka hasil evaluasi tidak dapat mencerminkan kemampuan peserta didik secara sebenarnya. Dengan demikian, validitas menjadi tolok ukur utama dalam menentukan kualitas instrumen evaluasi.

Dalam pengembangan instrumen pendidikan, validitas sering dikaitkan dengan kesesuaian antara butir instrumen dan ruang lingkup materi pembelajaran. Instrumen yang mencakup seluruh kompetensi yang diharapkan akan memberikan gambaran hasil belajar yang lebih utuh. Selain itu, validitas juga berkaitan dengan kejelasan konstruk yang diukur, sehingga instrumen tidak menimbulkan penafsiran ganda atau bias pengukuran. Ketepatan bahasa, struktur soal, serta kesesuaian konteks sangat memengaruhi tingkat validitas suatu instrumen evaluasi.

Selain validitas, reliabilitas merupakan aspek penting yang menentukan kualitas instrumen evaluasi. Reliabilitas berkaitan dengan konsistensi hasil pengukuran yang diperoleh dari suatu instrumen. Instrumen yang reliabel akan menghasilkan skor yang relatif stabil apabila digunakan dalam kondisi yang serupa, baik pada waktu yang berbeda maupun pada kelompok peserta didik

yang sebanding. Konsistensi ini menjadi dasar kepercayaan terhadap hasil evaluasi yang diperoleh.

Dalam praktik pendidikan, reliabilitas instrumen dipengaruhi oleh berbagai faktor, seperti kejelasan petunjuk pengerjaan, jumlah butir soal, tingkat kesukaran, serta objektivitas dalam proses penskoran. Instrumen dengan butir yang homogen dan terstruktur dengan baik cenderung menghasilkan tingkat reliabilitas yang lebih tinggi. Sebaliknya, instrumen yang ambigu atau memiliki penskoran yang subjektif berpotensi menurunkan konsistensi hasil pengukuran.

Hubungan antara validitas dan reliabilitas bersifat saling melengkapi.

Instrumen evaluasi yang berkualitas harus memenuhi kedua aspek tersebut secara bersamaan. Konsistensi pengukuran tanpa ketepatan sasaran tidak akan memberikan makna yang signifikan bagi evaluasi pembelajaran. Demikian pula, instrumen yang dirancang untuk mengukur kompetensi tertentu secara tepat tidak akan bernilai apabila hasil pengukurannya tidak konsisten. Oleh karena itu, validitas dan reliabilitas dipandang sebagai satu kesatuan yang tidak dapat dipisahkan dalam pengembangan instrumen evaluasi.

Dalam konteks pendidikan, khususnya pendidikan Islam, validitas dan reliabilitas instrumen evaluasi memiliki peran strategis dalam menjamin keadilan dan objektivitas penilaian. Instrumen evaluasi tidak hanya berfungsi untuk mengukur penguasaan materi, tetapi juga untuk menilai sikap, nilai, dan praktik keagamaan peserta didik. Hal ini menuntut instrumen yang tidak hanya akurat dan konsisten, tetapi juga sensitif terhadap karakteristik nilai-nilai pendidikan Islam.

Penerapan prinsip validitas dan reliabilitas dalam evaluasi pembelajaran memungkinkan pendidik memperoleh data yang dapat dijadikan dasar refleksi dan perbaikan proses pembelajaran. Evaluasi tidak lagi dipahami sekadar sebagai alat pengukur hasil belajar, melainkan sebagai bagian integral dari proses pembelajaran yang berkelanjutan. Dengan instrumen yang valid dan reliabel, evaluasi dapat berfungsi sebagai sarana peningkatan mutu pendidikan secara menyeluruh.

METHODOLOGY

Penulisan artikel ini menggunakan pendekatan *library research* atau studi kepustakaan, yaitu metode penulisan yang berfokus pada pengkajian dan analisis berbagai sumber literatur yang relevan dengan topik penulisan. Pendekatan ini dipilih karena tujuan penulisan adalah untuk membahas konsep, teori, serta kajian ilmiah yang berkaitan dengan validitas dan reliabilitas instrumen evaluasi secara mendalam dan konseptual. Melalui studi

kepuustakaan, peneliti dapat menghimpun berbagai pandangan ahli serta hasil penulisan terdahulu yang membahas pengukuran pendidikan, evaluasi hasil belajar, teknik pengujian instrumen, serta pengolahan skor hasil evaluasi dalam konteks pendidikan formal.

Sumber data dalam artikel ini terdiri atas buku teks akademik, jurnal ilmiah nasional dan internasional, prosiding seminar, serta dokumen ilmiah lainnya yang membahas validitas, reliabilitas, dan teknik evaluasi pembelajaran. Pengumpulan data dilakukan melalui proses penelusuran literatur dengan menggunakan kata kunci yang relevan, kemudian dilakukan seleksi sumber berdasarkan tingkat relevansi, kredibilitas, dan keterbaruan informasi. Literatur yang telah terpilih selanjutnya dibaca secara kritis untuk mengidentifikasi konsep utama, prinsip dasar, serta temuan penting yang berkaitan dengan fokus penulisan.

Analisis data dalam artikel ini dilakukan menggunakan pendekatan *descriptive qualitative*, dengan cara mengorganisasikan, mengelompokkan, dan mensintesis informasi dari berbagai sumber literatur. Data yang telah dianalisis kemudian disajikan dalam bentuk uraian naratif yang sistematis dan logis, sehingga mampu menggambarkan secara utuh konsep validitas dan reliabilitas instrumen evaluasi, contoh pengolahan skor mentah menjadi nilai huruf, pengubahan skor mentah menjadi skor standar *Z-score*, serta cara menguji dan meningkatkan kualitas instrumen evaluasi. Untuk menjaga keabsahan data, dilakukan perbandingan antar sumber literatur sehingga pembahasan yang dihasilkan bersifat objektif dan dapat dipertanggungjawabkan secara ilmiah.

RESULT AND DISCUSSION

Jenis Validitas Dan Reabilitas

Validitas dan reliabilitas merupakan dua konsep utama yang tidak dapat dipisahkan dalam proses evaluasi dan pengukuran, khususnya dalam bidang pendidikan. Keduanya berperan penting dalam menentukan kualitas suatu instrumen penilaian, karena berkaitan dengan ketepatan dan konsistensi hasil pengukuran. Validitas menunjukkan sejauh mana instrumen mampu mengukur apa yang seharusnya diukur sesuai dengan tujuan yang ditetapkan, sedangkan reliabilitas berkaitan dengan tingkat keajegan hasil yang diperoleh ketika pengukuran dilakukan secara berulang dalam kondisi yang relatif sama. Tanpa adanya validitas dan reliabilitas yang memadai, hasil pengukuran berpotensi menghasilkan data yang kurang akurat dan sulit dipercaya.

Selain sebagai indikator kualitas instrumen, validitas dan reliabilitas juga memiliki beragam jenis yang disesuaikan dengan kebutuhan dan konteks pengukuran. Setiap jenis validitas, seperti validitas isi, konstruk, dan kriteria, memiliki pendekatan yang berbeda dalam memastikan ketepatan pengukuran. Demikian pula, reliabilitas dapat ditinjau dari berbagai aspek, seperti konsistensi waktu, kesesuaian antarpemilai, dan keterkaitan antarbutir instrumen. Pemahaman terhadap jenis-jenis validitas dan reliabilitas ini

menjadi dasar penting bagi peneliti dan pendidik dalam menyusun serta menggunakan instrumen penilaian secara tepat, sehingga hasil yang diperoleh lebih bermakna dan dapat dipertanggungjawabkan.

Validitas

Istilah validitas berasal dari bahasa Latin *validus* yang bermakna kuat atau kokoh. Secara konseptual, validitas dapat dibedakan menjadi dua jenis, yaitu validitas penelitian dan validitas pengukuran. Validitas penulisan berkaitan dengan tingkat kebenaran kesimpulan yang dihasilkan dari suatu penulisan, yang sangat dipengaruhi oleh ketepatan metode penulisan, representativitas sampel, serta karakteristik populasi yang menjadi dasar pengambilan sampel (Zayrin et al., 2025). Sebagai ilustrasi, kesimpulan dalam sebuah meta-analisis yang melibatkan sejumlah penelitian dan menyatakan adanya peningkatan risiko tumor otak akibat penggunaan telepon seluler dalam jangka panjang mencerminkan aspek validitas penulisan. Namun demikian, pembahasan dalam kajian ini difokuskan pada validitas pengukuran, bukan pada validitas penulisan.

Validitas pengukuran mengacu pada tingkat ketepatan dan kecermatan suatu instrumen dalam menjalankan fungsinya sebagai alat ukur. Validitas juga dipahami sebagai kelayakan interpretasi dan pemanfaatan hasil penilaian yang dihasilkan oleh instrumen tersebut (Fadli et al., 2023). Dengan demikian, validitas menunjukkan sejauh mana suatu alat ukur mampu mengukur aspek yang memang dimaksudkan untuk diukur. Dalam konteks instrumen pengumpulan data, seperti kuesioner, validitas tercermin dari keterkaitan yang kuat antarbutir pertanyaan yang diukur melalui korelasi jawaban responden; butir pertanyaan yang memiliki korelasi rendah dinilai kurang valid. Selain itu, validitas juga mencakup kesesuaian instrumen dengan landasan teori yang digunakan serta adanya konsistensi internal antarbagian instrumen, sehingga produk yang dihasilkan benar-benar layak digunakan sebagai alat ukur (Anshari et al., 2024).

Jenis-jenis Validitas Tes

Secara umum, ada dua konsep utama validitas dalam evaluasi instrumen, yaitu validitas tes dan validitas butir soal. Validitas tes mencakup validitas logis dan validitas empiris, di mana validitas logis berkaitan dengan sejauh mana sebuah instrumen dinyatakan valid berdasarkan alasan teoritis dan rancangan yang sesuai dengan tolok ukur penilaian, sedangkan validitas empiris diuji melalui bukti statistik dari hasil pelaksanaan tes. Dalam validitas logis itu sendiri terdapat dua bentuk penting, yaitu validitas isi yang menilai kesesuaian butir-butir soal dengan tujuan atau materi pengukuran, serta validitas konstruk yang menilai sejauh mana skor yang dihasilkan instrumen mencerminkan konsep teoretis yang mendasarinya. Studi tentang validitas tes dan kualitas butir soal menunjukkan bahwa validitas tes harus mempertimbangkan konstruksi teoretis dan hubungan antar-butir agar instrumen benar-benar menggambarkan apa yang ingin diukur secara menyeluruh (Muzaffar, 2016).

Validitas isi menekankan pada representasi materi evaluasi yang sesuai dengan tujuan pembelajaran dan kurikulum, sedangkan validitas konstruk menekankan pada

kesesuaian pengukuran terhadap konsep teoretis yang ingin diungkap melalui instrumen tersebut. Kedua jenis validitas logis ini merupakan landasan awal dalam menilai kualitas instrumen sebelum dilakukan pengujian empiris yang lebih lanjut. Penulisan nasional yang menganalisis validitas isi dan validitas konstruk pada soal Ujian Akhir Semester Bahasa Indonesia menunjukkan bahwa validitas isi berkaitan dengan ketercakupan indikator pembelajaran yang ditetapkan, sedangkan validitas konstruk berhubungan dengan kemampuan soal dalam mengukur kompetensi yang diharapkan sesuai dengan definisi konsepnya (Sihotang et al., 2025).

Faktor yang Mempengaruhi Validitas

Terdapat tiga faktor yang memengaruhi validitas hasil tes, yakni (Haq, 2022) :

a. Faktor instrumen evaluasi

Membuat instrumen evaluasi merupakan tugas yang kompleks, terutama jika evaluator tidak memahami sepenuhnya prosedur dan teknik evaluasi. Jika instrumen evaluasi tidak disusun dengan baik, hasil evaluasi juga dapat terpengaruh secara negatif. Oleh karena itu, dalam proses pengembangan instrumen evaluasi, seorang evaluator perlu memperhatikan faktor-faktor yang memengaruhi validitas instrumen dan terkait dengan prosedur penyusunan instrument.

b. Faktor administrasi evaluasi dan penskoran

Dalam pelaksanaan evaluasi dan proses penilaian, seringkali terdapat berbagai penyimpangan atau kesalahan, seperti distribusi waktu yang tidak seimbang dalam mengerjakan soal, memberikan bantuan kepada peserta didik dengan metode yang beragam, kesalahan dalam proses penilaian, dan kondisi fisik serta mental peserta didik yang tidak mendukung.

c. Faktor jawaban dari peserta didik

Secara praktis, faktor yang memiliki dampak lebih besar adalah respons peserta didik, lebih dominan dibandingkan dengan dua faktor sebelumnya. Faktor ini mencakup kebiasaan peserta didik dalam memberikan jawaban dengan cepat namun tidak tepat, serta keinginan untuk mencoba-coba.

Teknik Menentukan Validitas Tes

Dalam mengevaluasi validitas alat ukur, yang mengukur sejauh mana alat tersebut mencapai tujuan pengukuran, secara umum, pendekatan konvensional melibatkan tiga aspek, yakni: a) isi yang diukur, b) rekaan teoritis (konstruk) dari atribut yang diukur, dan c) kriteria yang digunakan oleh alat ukur. Oleh karena itu, dalam literatur tentang validitas tes, terdapat tiga jenis validitas, yakni validitas isi, validitas konstruk, dan validitas kriteria (Bashooir & Supahar, 2018) (Sihotang et al., 2025).

a. Validitas isi

Validitas isi, yang sering juga disebut sebagai validitas kurikulum, mencerminkan bahwa suatu alat ukur dianggap valid jika sesuai dengan materi kurikulum yang hendak diukur. Validitas isi tes menekankan pada kualitas dan relevansi pertanyaan atau item tes dengan isi kurikulum yang terkait. Salah satu

metode untuk menilai validitas isi adalah melalui pemeriksaan terhadap isi tes itu sendiri, memastikan kesesuaian dengan materi yang diinginkan.

b. Validitas konstruk

Construct validity adalah bentuk validitas yang mencerminkan sejauh mana suatu tes dapat mengukur suatu ciri atau konstruk teoritis (sering disebut sebagai variabel laten) yang ingin diukur.

c. Validitas kriteria

Validitas berdasarkan kriteria dapat dibagi menjadi dua jenis, tergantung pada kapan kriteria tersebut dapat digunakan. Jika kriteria tersebut dapat digunakan sekarang atau dalam waktu dekat, disebut sebagai validitas sejalan (concurrent validity), sementara jika kriteria tersebut baru dapat dimanfaatkan beberapa waktu ke depan, disebut sebagai validitas prediktif (predictive validity).

Reliabilitas

Reliabilitas instrumen berkaitan dengan tingkat konsistensi dan kestabilan hasil pengukuran yang dihasilkan oleh suatu alat ukur. Instrumen dikatakan reliabel apabila mampu memberikan hasil yang relatif sama ketika digunakan berulang kali pada kondisi yang serupa, sehingga skor yang diperoleh dapat dipercaya sebagai representasi kemampuan atau karakteristik yang diukur. Dalam evaluasi pendidikan, reliabilitas penting karena menunjukkan sejauh mana instrumen mampu meminimalkan kesalahan pengukuran dan menghasilkan data yang stabil, yang umumnya diuji melalui konsistensi internal, seperti koefisien Cronbach's Alpha (Isma & Yusuf, 2025).

Selain itu, reliabilitas juga mencerminkan keandalan hasil ukur ketika pengukuran dilakukan oleh penguji yang berbeda, pada waktu yang berbeda, atau pada kelompok subjek yang sebanding. Instrumen yang tidak reliabel akan menghasilkan skor yang tidak konsisten sehingga dapat menurunkan kualitas interpretasi dan kesimpulan penulisan. Oleh karena itu, reliabilitas merupakan prasyarat penting dalam pengukuran pendidikan dan penulisan sosial, karena tanpa reliabilitas yang memadai, validitas hasil pengukuran juga sulit dicapai (Bashooir & Supahar, 2018).

Manfaat Reliabilitas

Reliabilitas memiliki peran penting dalam penilaian pembelajaran karena menentukan kualitas dan kebermaknaan hasil evaluasi yang diperoleh. Evaluasi pembelajaran memberikan berbagai manfaat, antara lain membantu memahami kondisi pembelajaran secara menyeluruh, seperti karakteristik awal peserta didik, motivasi belajar, ketersediaan sarana dan prasarana, serta situasi pembelajaran yang melibatkan peserta didik dan pendidik. Selain itu, evaluasi berfungsi sebagai dasar dalam pengambilan keputusan terkait keberlanjutan program pembelajaran dan penanganan berbagai permasalahan yang muncul, serta sebagai upaya untuk meningkatkan kualitas proses belajar mengajar melalui perbaikan komponen-komponen pembelajaran. Secara khusus, evaluasi pembelajaran juga memberikan manfaat langsung bagi pihak-pihak yang terlibat, seperti peserta didik, guru, dan pimpinan sekolah; bagi peserta didik, evaluasi

membantu menilai tingkat pencapaian tujuan pembelajaran dan menentukan apakah hasil yang diperoleh sudah memuaskan atau masih perlu ditingkatkan (Siregar, 2018).

Metode estimasi reliabilitas

Allen dan Yen dalam Khumaedi, mengemukakan bahwa terdapat tiga pendekatan utama dalam mengestimasi koefisien reliabilitas, yaitu metode tes ulang (test–retest), metode bentuk paralel (parallel forms), dan metode konsistensi internal (internal consistency). Ketiga metode tersebut digunakan sesuai dengan karakteristik instrumen dan tujuan pengukuran yang ingin dicapai (Khumaedi, 2012).

Metode tes ulang dilakukan dengan cara memberikan instrumen yang sama kepada kelompok responden yang sama dalam dua waktu pelaksanaan yang berbeda. Hubungan antara skor hasil pengukuran pertama dan kedua kemudian dianalisis untuk memperoleh koefisien reliabilitas. Jarak waktu antar pelaksanaan tes bergantung pada tujuan penggunaan instrumen. Instrumen yang digunakan untuk keperluan peramalan jangka panjang, seperti keberhasilan kerja, dapat memiliki interval waktu hingga beberapa tahun, sedangkan instrumen yang mengukur hasil belajar atau motivasi umumnya menggunakan interval yang lebih singkat, sekitar dua hingga tiga minggu. Namun demikian, metode ini berpotensi menimbulkan efek pengulangan, di mana hasil pengukuran kedua dapat dipengaruhi oleh pengalaman pada pengukuran pertama. Selain itu, adanya proses belajar atau perolehan informasi baru oleh responden di antara dua waktu pengukuran dapat menyebabkan perubahan skor yang berdampak pada besarnya koefisien reliabilitas.

Tabel 3.1 Cara menentukan Koefisien reliabilitas metode estimasi tes ulang

Sampel	X	Y	X ²	Y ²	X.Y
1	X ₁	Y ₁	X ₁ ²	Y ₁ ²	X ₁ .Y ₁
2	X ₂	Y ₂	X ₂ ²	Y ₂ ²	X ₂ .Y ₂
n	X _n	Y _n	X _n ²	Y _n ²	X _n .Y _n
Σ	X	Y	X ²	Y ²	X.Y

Rumus korelasi product moment digunakan untuk memasukkan data tersebut:

$$r_{xy} = \frac{n \cdot \sum X.Y - (\sum X) \cdot (\sum Y)}{\sqrt{\{n \cdot \sum X^2 - (\sum X)^2\} \cdot \{n \cdot \sum Y^2 - (\sum Y)^2\}}}$$

Metode bentuk paralel menggunakan dua instrumen yang dirancang setara dari segi materi, aspek yang diukur, serta tingkat kesulitan. Kedua instrumen tersebut dapat diberikan secara bersamaan kepada responden. Tantangan utama dalam metode ini terletak pada penyusunan dua instrumen yang benar-benar paralel, sehingga memerlukan ketelitian tinggi dalam proses pengembangannya. Instrumen dikatakan paralel apabila memiliki rerata, varians, dan distribusi skor yang relatif sama serta menunjukkan korelasi yang tinggi. Setelah pengujian dilakukan, skor dari instrumen pertama dan instrumen paralel dianalisis menggunakan korelasi Pearson untuk memperoleh koefisien reliabilitas.

Tabel 3.2 Cara menentukan Koefisien reliabilitas metode bentuk paralel

Sampel	X	$x = X - \bar{x}$	Y	$y = Y - \bar{y}$	X^2	Y^2	$X.Y$
1	X_1	x_1	Y_1	y_1	X_1^2	Y_1^2	$X_1.Y_1$
2	X_2	x_2	Y_2	y_2	X_2^2	Y_2^2	$X_2.Y_2$
n	X_n	x_n	Y_n	y_n	X_n^2	Y_n^2	$X_n.Y_n$
Σ	X	x	Y	y	X^2	Y^2	$X.Y$

Rumus korelasi pearson digunakan untuk memasukkan data tersebut:

$$r_{xy} = \frac{\sum x.y}{\sqrt{(\sum x^2 . \sum y^2)}}$$

Metode konsistensi internal hanya menggunakan satu instrumen dan dilakukan dalam satu kali pengujian. Pendekatan ini menilai reliabilitas berdasarkan konsistensi antarbutir dalam instrumen dengan cara membagi instrumen ke dalam beberapa bagian yang setara. Pembagian tersebut diupayakan menghasilkan belahan yang paralel agar estimasi reliabilitas lebih akurat. Metode konsistensi internal merupakan pendekatan yang paling sering digunakan karena tidak memerlukan pengulangan pengukuran, sehingga berbagai permasalahan akibat pengulangan dapat dihindari. Perhitungan reliabilitas dengan metode ini dapat dilakukan melalui beberapa rumus, di antaranya rumus Spearman-Brown, Alpha Cronbach, dan Kuder-Richardson, yang masing-masing disesuaikan dengan karakteristik data dan jenis instrumen yang digunakan.

Tabel 3.3 Cara menentukan korelasi antara skor belahan ganjil dan genap

Sampel	X	Y	X^2	Y^2	$X.Y$
1	X_1	Y_1	X_1^2	Y_1^2	$X_1.Y_1$
2	X_2	Y_2	X_2^2	Y_2^2	$X_2.Y_2$
n	X_n	Y_n	X_n^2	Y_n^2	$X_n.Y_n$
Σ	X	Y	X^2	Y^2	$X.Y$

Rumus reliabilitas spearman brown digunakan untuk memasukkan data tersebut:

$$r_{xy} = \frac{2r_{xy}}{1 + r_{xy}}$$

Contoh Pengolahan Skor Mentah Menjadi Nilai Huruf

Konversi skor mentah menjadi nilai huruf adalah proses transformasi skor yang diperoleh peserta didik dalam bentuk angka mentah (raw score) menjadi kategori nilai yang lebih mudah diinterpretasikan (misalnya A, B, C, D, E atau skala lainnya). Langkah umum pengolahan ini melibatkan:(Damayanti et al., 2023)

- Menghitung mean (rata-rata) dan standar deviasi dari skor mentah seluruh responden.
- Menetapkan batas nilai huruf berdasarkan interval nilai yang ditentukan (mis. PAP atau PAN).
- Menentukan nilai huruf untuk setiap peserta dengan membandingkan skor mentah mereka terhadap interval tersebut.

Metode ini sering dipakai dalam penulisan pendidikan untuk mempermudah interpretasi hasil tes atau menilai pencapaian kompetensi peserta didik

Mengolah Skor Mentah Menjadi Skor Standart Z

Skor Z (standard score) adalah skor yang menunjukkan posisi sebuah skor mentah relatif terhadap rata-rata kelompok dalam satuan standar deviasi. Rumusnya adalah: (Damayanti et al., 2023)

$$Z = \frac{X - \bar{X}}{s}$$

di mana X adalah skor mentah, $\bar{X}$ adalah rata-rata, dan s adalah standar deviasi.

Transformasi ini memudahkan perbandingan antar peserta maupun antar instrumen, karena skor Z menunjukkan seberapa jauh skor individu berada di atas atau di bawah rata-rata populasi. Skor Z sering dipakai dalam penulisan pendidikan untuk standarisasi data, normalisasi distribusi skor, dan untuk mengurangi bias akibat perbedaan skala atau keragaman soal.

Cara Menguji dan Meningkatkan Keduanya

Untuk menguji kualitas pengolahan skor baik dalam bentuk nilai huruf maupun skor Z, berikut pendekatan yang sering digunakan dalam literatur ilmiah: (Sumin et al., 2022)

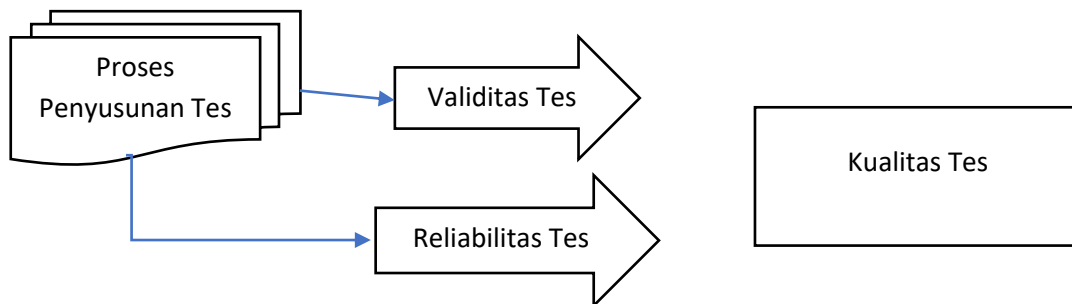
- a. Uji Validitas
 - 1) Validitas isi: pemeriksaan oleh pakar apakah skor atau konversinya sesuai dengan indikator yang diukur.
 - 2) Validitas konstruk: konfirmasi melalui analisis faktor atau model psikometrika apakah skor memang mencerminkan konstruk teoretis yang diukur.
 - 3) Transformasi skor seperti Z-score dianalisis untuk melihat apakah distribusi skor hasil transformasi tetap mencerminkan struktur konstruk yang valid.
- b. Uji Reliabilitas
 - 1) Konsistensi internal: Cronbach's Alpha atau metode lain digunakan untuk menilai apakah item-item setelah transformasi skala tetap konsisten.
 - 2) Transformasi (seperti Z-score) tidak otomatis meningkatkan reliabilitas, namun dapat membantu menstandarisasi data sehingga analisis reliabilitas menjadi lebih stabil secara statistik. Studi yang sama menyimpulkan bahwa rescaling dengan Z-score dapat meningkatkan validitas faktor tetapi tidak selalu meningkatkan nilai reliabilitas internal

Pengujian Ulang Validitas dan Reliabilitas Instrumen Baku

Pengujian ulang validitas dan reliabilitas pada instrumen baku pada dasarnya dapat mencakup seluruh jenis pengujian yang lazim diterapkan pada instrumen baru, seperti validitas isi, validitas konstruk, dan validitas kriteria, namun pelaksanaannya dapat disesuaikan dengan tujuan penulisan serta konteks penggunaan instrumen (Efsthathiou, 2018). Apabila instrumen digunakan untuk mengukur variabel yang sama pada populasi dengan karakteristik yang relatif serupa dengan penulisan sebelumnya, pengujian ulang terhadap validitas kriteria dan reliabilitas konsistensi internal umumnya dianggap memadai. Validitas kriteria dapat diuji melalui analisis korelasi antara skor instrumen dengan kriteria eksternal yang relevan, sedangkan validitas konstruk perlu diuji kembali apabila instrumen diterapkan untuk mengukur konstruk atau landasan teori yang

berbeda atau ketika terjadi perubahan definisi variabel yang diukur. Teori yang mendasari konstruk berperan sebagai acuan dalam perancangan dan evaluasi instrumen, sementara reliabilitas biasanya diuji melalui konsistensi internal menggunakan koefisien Cronbach's Alpha untuk menilai tingkat keajegan antarbutir dalam skala. Pengujian ulang tersebut harus melibatkan sampel yang representatif terhadap populasi sasaran, baik dari populasi utama maupun populasi dengan karakteristik serupa, dengan tetap memperhatikan kehati-hatian agar tidak terjadi bias akibat penggunaan responden yang sama secara berulang.

Analisis Validitas dan Reliabilitas pada Instrumen Evaluasi



Gambar 3.1. Kerangka berfikir kualitas sebuah tes

Berdasarkan bagan tersebut, dapat disimpulkan bahwa kualitas suatu tes diawali dengan penilaian terhadap proses penyusunannya melalui analisis kualitatif. Tahap berikutnya adalah melakukan analisis kuantitatif yang berfokus pada dua aspek utama, yaitu validitas dan reliabilitas tes. Hasil dari analisis tersebut kemudian diinterpretasikan, dan interpretasi yang diperoleh digunakan sebagai dasar atau tolok ukur dalam menentukan mutu tes (Khafidin, 2014).

Hasil penulisan menunjukkan bahwa validitas dan reliabilitas merupakan aspek utama yang menentukan mutu instrumen evaluasi yang digunakan. Instrumen yang dianalisis telah disusun berdasarkan landasan teoretis yang jelas serta tujuan pengukuran yang terdefinisi dengan baik, sehingga mampu merepresentasikan konstruk yang ingin diukur. Hal ini menegaskan bahwa kualitas hasil pengukuran sangat dipengaruhi oleh ketepatan perencanaan dan penyusunan instrumen sejak tahap awal.

Dari sisi validitas, hasil analisis menunjukkan bahwa instrumen memenuhi kriteria validitas isi dan validitas konstruk. Validitas isi terlihat dari kesesuaian butir-butir instrumen dengan indikator dan materi yang diukur, sedangkan validitas konstruk tercermin dari kemampuan instrumen dalam menggambarkan konsep teoretis yang mendasari variabel penulisan. Temuan ini menunjukkan bahwa instrumen secara konseptual layak digunakan sebelum dilakukan pengujian lebih lanjut secara empiris.

Selain itu, validitas kriteria turut memperkuat keakuratan hasil pengukuran dengan menunjukkan adanya keterkaitan antara skor instrumen dan kriteria eksternal yang relevan. Hasil penulisan mengindikasikan bahwa instrumen memiliki kemampuan yang memadai baik untuk menilai kondisi saat ini maupun untuk memprediksi capaian di

masa mendatang. Dengan demikian, hasil pengukuran yang diperoleh dapat digunakan sebagai dasar pengambilan keputusan yang lebih tepat.

Ditinjau dari aspek reliabilitas, hasil penulisan menunjukkan bahwa instrumen memiliki tingkat konsistensi yang baik. Pengujian reliabilitas melalui pendekatan konsistensi internal memperlihatkan adanya keterkaitan yang stabil antarbutir instrumen, sehingga skor yang dihasilkan relatif ajeg. Temuan ini menunjukkan bahwa kesalahan pengukuran dapat diminimalkan dan hasil tes dapat dipercaya untuk menggambarkan kemampuan atau karakteristik yang diukur.

Hasil penulisan juga menunjukkan bahwa pengolahan skor menjadi bagian penting dalam interpretasi hasil evaluasi. Transformasi skor mentah ke dalam nilai huruf membantu menyederhanakan hasil pengukuran dan memudahkan pengelompokan tingkat pencapaian peserta didik. Proses ini menjadikan hasil evaluasi lebih komunikatif dan mudah dipahami oleh berbagai pihak yang berkepentingan. Selain nilai huruf, penggunaan skor standar Z memberikan gambaran yang lebih objektif mengenai posisi capaian peserta didik dalam kelompok. Skor Z memungkinkan perbandingan hasil secara lebih proporsional karena menunjukkan jarak skor individu terhadap rata-rata kelompok dalam satuan standar deviasi. Hasil penulisan menunjukkan bahwa transformasi ini sangat membantu dalam analisis statistik lanjutan tanpa mengubah makna konstruk yang diukur.

Secara keseluruhan, hasil penulisan menegaskan bahwa kualitas evaluasi tidak hanya ditentukan oleh validitas dan reliabilitas instrumen, tetapi juga oleh cara hasil pengukuran diolah dan disajikan. Instrumen yang valid dan reliabel perlu didukung oleh teknik pengolahan skor yang tepat agar hasil evaluasi benar-benar bermakna dan dapat dipertanggungjawabkan. Dengan demikian, validitas, reliabilitas, dan transformasi skor merupakan satu kesatuan yang saling melengkapi dalam menghasilkan evaluasi yang berkualitas.

CONCLUSION AND RECOMMENDATION

Artikel ini memiliki keterkaitan yang kuat serta berpotensi menjadi referensi bagi penulisan lain dalam kajian analisis instrumen evaluasi. Hal tersebut didukung oleh beberapa alasan. Pertama, dari sisi kontribusi metodologis, penulisan ini memberikan pemahaman yang komprehensif mengenai proses pengembangan instrumen evaluasi yang memenuhi kriteria valid dan reliabel, sehingga dapat membantu peneliti lain dalam merancang penulisan dengan pendekatan metodologi yang sesuai. Kedua, dari aspek penerapan praktis, hasil penulisan dapat menjadi pedoman bagi praktisi pendidikan, guru, maupun pengembang kurikulum dalam memilih atau menyusun instrumen evaluasi yang tepat untuk menilai pemahaman peserta didik. Ketiga, penulisan ini juga membuka peluang perluasan konteks, karena temuan yang dihasilkan dapat diterapkan atau dikaji lebih lanjut dalam berbagai konteks pendidikan lainnya. Keempat, penulisan ini dapat dijadikan sebagai acuan dalam penulisan lanjutan yang bertujuan untuk membandingkan atau mereplikasi temuan dengan konteks maupun metode

yang berbeda. Dengan demikian, kajian mengenai pentingnya validitas dan reliabilitas instrumen evaluasi memiliki implikasi dan kontribusi yang luas dalam pengembangan bidang pendidikan secara umum

REFERENCES

- Anshari, M. I., Nasution, R., Irsyad, M., Alifa, A. Z., & Zuhriyah, I. A. (2024). Analisis Validitas dan Reliabilitas Butir Soal Sumatif Akhir Semester Ganjil Mata Pelajaran PAI. *Edukatif: Jurnal Ilmu Pendidikan*, 6(1), 964–975. <https://doi.org/10.31004/edukatif.v6i1.5931>
- Bashooir, K., & Supahar. (2018). Validity and reability instrument of scientific literacy performance assesment in physics teaching based on STEM. *Jurnal Penulisan Dan Evaluasi Pendidikan*, 22(2), 168–181. <https://doi.org/doi.org/10.21831/pep.v22i2.19590>
- Damayanti, A. M., Daryono, & Rayanto, Y. H. (2023). *Evaluasi Pembelajaran* (Vol. 16, Issue 1). Basya Media Utama.
- Efstathiou, G. (2018). Translation, Adaptation and Validation Process of Research Instruments. *Spring International Publishing*. https://doi.org/doi.org/10.1007/978-3-319-89899-5_7
- Fadli, R., Hidayati, S., Cholifah, M., Siroj, R. A., & Afgani, M. W. (2023). Validitas dan Reliabilitas pada Penulisan Motivasi Belajar Pendidikan Agama Islam Menggunakan Product Moment. *JlIP - Jurnal Ilmiah Ilmu Pendidikan*, 6(3), 1734–1739. <https://doi.org/10.54371/jiip.v6i3.1419>
- Haq, V. A. (2022). Menguji Validitas Dan Reliabilitas Pada Mata Pelajaran Al Qur'an Hadits Menggunakan Korelasi Produk Momenspearman Brown. *An-Nawa : Jurnal Studi Islam*, 4(1), 11–24. <https://doi.org/10.37758/annawa.v4i1.419>
- Isma, N., & Yusuf, M. (2025). *The Influence of the Implementation of Extracurricular Activities of the Islamic Propagation Agency on the Practice of Religious Worship at Mutia Rahma Bulu Cina Middle School , Hamparan Perak District*. 5(1), 211–215. <https://doi.org/10.30596/jcositte.v1i1.xxxx>
- Khafidin, Z. (2014). Analisis Validitas Dan Reliabilitas Tes Mata Pelajaran Pendidikan Agama Islam Tingkat Sma. *Edukasia: Jurnal Penulisan*

Pendidikan Islam, 9(2), 253–266.

<https://doi.org/10.21043/edukasia.v9i2.775>

Khumaedi, M. (2012). The reliability of education research instruments. In *Jurnal Pendidikan Teknik Mesin* (Vol. 12, Issue 1, pp. 25–30).

<https://journal.unnes.ac.id/nju/index.php/JPTM/article/view/5273>

Muzaffar, A. (2016). Validitas Tes dan Kualitas Butir Soal. *LISANUNA: Jurnal Ilmu Bahasa Arab Dan Pembelajarannya*, 5(1).

<https://doi.org/doi.org/10.22373/1.v5i1.859>

Rachman, A., Yochanan, E., Samanlangi, andi ilham, & Purnomo, H. (2024). *Metode Penulisan Kuantitatif, Kualitatif, dan R&D*. CV Saba jaya publisher.

Sihotang, A. E., Hutagalung, S. H., Saragih, S., & Akbar, R. F. (2025). JKIP : Jurnal Kajian Ilmu Pendidikan Content and Construct Validity Analysis of Indonesian Language Final Exam Questions for Class VIII of SMP Harapan 1 Medan Analisis Validitas Isi Dan Konstruk Pada Soal Ujian Akhir Semester Bahasa Indonesia Kelas VIII S. *JKIP: Jurnal Kajian Ilmu Pendidikan*, 6(2), 536–544.

<https://doi.org/doi.org/10.55583/jkip.v6i2.1385>

Siregar, raja lottung. (2018). Evaluasi Hasil Belajar. *HIKMAH: Jurnal Pendidikan Islam*, 6(1), 59–75. <https://doi.org/dx.doi.org/10.55403/hikmah.v6i1.40>

Sumin, Sukmawati, F., Setiawati, F. A., & Asmawi, S. (2022). The Impact of Z-Score Transformation Scaling on the Validity, Reliability, and Measurement Error of Instrument SATS-36. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 11(2), 166–180.

<https://doi.org/10.15408/jp3i.v11i2.26591>

Zayrin, afifah aulia, Nupus, H., Maizia, K. khansa, Marsela, S., Hidatullah, R., & Harmonedi. (2025). Analisis Instrumen Penulisan Pendidikan (Uji Validitas Dan Relibilitas Instrumen Penulisan). *Jurnal QOSIM Jurnal Pendidikan Sosial & Humaniora*, 3(2), 780–789.

<https://doi.org/10.61104/jq.v3i2.1070>