



VALIDITAS DAN RELIABILITAS DALAM EVALUASI PAI

Faiqotun Nisa¹, Bashiratud Diyana Basy², Ahmad Jauhari³, Yahya Mof⁴, Hasni Noor⁵
Universitas Islam Negeri Antasari, Banjarmasin

faiqohannisa16@gmail.com¹, bashiratuddiyana20@gmail.com², ahmadjauhari6@gmail.com³,
mofyahya@gmail.com⁴, Hasninoor@uin-antasari.ac.id⁵

Abstrak

Validitas dan reliabilitas merupakan dua aspek fundamental yang menentukan kualitas instrumen evaluasi pembelajaran. Instrumen yang valid mampu mengukur kompetensi yang seharusnya diukur, sedangkan instrumen yang reliabel menghasilkan data yang konsisten dan dapat dipercaya. Penelitian ini bertujuan untuk mengkaji konsep validitas dan reliabilitas instrumen evaluasi pembelajaran serta implikasinya dalam Pendidikan Agama Islam (PAI). Penelitian menggunakan pendekatan kualitatif dengan jenis penelitian kepustakaan (library research). Data diperoleh dari berbagai sumber primer dan sekunder berupa buku, jurnal ilmiah, serta dokumen akademik yang relevan. Analisis data dilakukan menggunakan teknik content analysis melalui tahapan reduksi data, penyajian data, interpretasi, dan penarikan kesimpulan. Hasil kajian menunjukkan bahwa validitas instrumen meliputi validitas isi, validitas konstruk, dan validitas kriteria yang berfungsi menjamin ketepatan pengukuran. Sementara itu, reliabilitas berkaitan dengan konsistensi hasil pengukuran yang dapat diuji melalui teknik test-retest, split-half, Cronbach Alpha, serta Kuder-Richardson (KR-20 dan KR-21). Temuan penelitian juga menunjukkan bahwa kualitas instrumen evaluasi dipengaruhi oleh jumlah butir soal, tingkat kesukaran, objektivitas penskoran, homogenitas konstruk, dan sebaran skor peserta didik. Oleh karena itu, pengujian validitas dan reliabilitas perlu dilakukan secara sistematis untuk menghasilkan instrumen evaluasi yang akurat, objektif, dan dapat dipertanggungjawabkan dalam proses pembelajaran Pendidikan Agama Islam.

Kata kunci: evaluasi pembelajaran; instrumen penilaian; pendidikan agama Islam; reliabilitas; validitas.

Abstract

Validity and reliability are two fundamental aspects that determine the quality of educational evaluation instruments. A valid instrument accurately measures the intended competencies, while a reliable instrument produces consistent and trustworthy results. This study aims to examine the concepts of validity and reliability of educational evaluation instruments and their implications for Islamic Religious Education (IRE). The study employed a qualitative approach using a library research method. Data were collected from primary and secondary sources, including books, scientific journals, and other relevant academic publications. Data analysis was conducted using content analysis through the stages of data reduction, data presentation, interpretation, and conclusion drawing. The findings indicate that instrument validity consists of content validity, construct validity, and criterion-related validity, all of which ensure measurement accuracy. Meanwhile, reliability refers to the consistency of measurement results and can be assessed through test-retest, split-half, Cronbach's Alpha, and Kuder-Richardson (KR-20 and KR-21) techniques. The study also reveals that the quality of evaluation instruments is influenced by factors such as the number of test items, difficulty level, scoring objectivity, construct homogeneity, and score

distribution. Therefore, validity and reliability testing should be systematically implemented to produce accurate, objective, and accountable evaluation instruments in the learning process of Islamic Religious Education.

Keywords: *educational evaluation; assessment instrument; Islamic religious education; reliability; validity.*

PENDAHULUAN

Pendidikan Agama Islam (PAI) merupakan salah satu mata pelajaran yang tidak sekadar proses transfer pengetahuan (*transfer of knowledge*). Pendidikan Agama Islam memiliki peran penting dalam membentuk karakter dan moral peserta didik¹. Sebagaimana inti pembelajarannya, PAI memiliki tanggung jawab besar dalam menanamkan nilai-nilai akidah, syariat, dan akhlak yang bersumber langsung dari Al-Qur'an dan As-Sunnah². Oleh karena itu, dalam setiap aktivitas pembelajaran, memerlukan sistem evaluasi yang akurat guna mengukur sejauh mana tujuan-tujuan mulia tersebut telah tercapai.

Sistem evaluasi pendidikan memiliki komponen penting untuk mengukur ketercapaian tujuan pembelajaran secara holistik, meliputi aspek kognitif, afektif, dan psikomotorik³. Dalam tahapannya, evaluasi terkadang menghadapi tantangan besar yang sering muncul. Sebuah instrumen penilaian atau tes yang diberikan kepada siswa tidak akan memiliki makna jika tidak memiliki validitas yang kuat. Aspek validitas ini menjadi fokus utama guna menjamin bahwa alat evaluasi yang digunakan benar-benar mengukur apa yang seharusnya diukur dan juga memberikan hasil yang dapat dipercaya⁴. Hal ini menjadi tantangan tersendiri bagi pendidik dalam mengukur hasil belajar menggunakan instrument yang valid.

Validitas instrument evaluasi berkaitan erat dengan sejauh mana instrument tersebut mampu mengukur apa yang diinginkan atau diukur⁵. Validitas menjadi ukuran ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurnya. Dalam konteks PAI, validitas menjadi sangat penting karena materi PAI yang diukur mengandung aspek spiritual dan normatif⁶. Validitas instrumen evaluasi PAI akan menentukan sejauh mana instrument tersebut mampu mencerminkan pemahaman, nilai-nilai, dan kompetensi keagamaan yang diinginkan oleh kurikulum atau standar pembelajaran yang berlaku.

¹ Derri Adi Fernando, Hartatiana Hartatiana, and Fajri Ismail, "Pentingnya Validitas Dan Reliabilitas Instrument Evaluasi Pada Pembelajaran Pendidikan Agama Islam," *Raudhah Proud To Be Professionals: Jurnal Tarbiyah Islamiyah* 8, no. 3 (2023): 1110–21.

² Oktio Frenki Biantoro and Asep Rahmatullah, "Internalisasi Nilai-Nilai Pendidikan Agama Islam Dalam Pembinaan Moral Siswa Di Sekolah," *Pelita: Jurnal Studi Islam Mahasiswa UII Dalwa* 2, no. 2 (2025): 225–41.

³ Laila Laila, Alawiyah Nabila, and Eka Widyanti, "Konsep Dasar Evaluasi Pembelajaran," *Jurnal Manajemen Dan Pendidikan Agama Islam* 2, no. 5 (2024): 252–62.

⁴ Anri Saputra and others, "Pengembangan Instrumen Evaluasi," *Ar-Raudah: Jurnal Pendidikan Dan Keagamaan* 2, no. 4 (2025): 1–14.

⁵ I Wayan Widiani et al., *Validasi Penyusunan Instrumen Penelitian Pendidikan* (PT. RajaGrafindo Persada-Rajawali Pers, 2023).

⁶ Wawan Hendrawan, Rudi Haryadi, and Cucu Atikah, "Instrumen Penilaian Pendidikan Agama Islam," *JTPPm (Jurnal Teknologi Pendidikan Dan Pembelajaran): Edutech and Intructional Research Journal* 11, no. 2 (2024).

Instrument evaluasi dalam PAI juga berkaitan dengan upaya peningkatan kualitas Pendidikan keagamaan⁷, yang mana hasil dari evaluasi yang akurat akan memberikan informasi yang berguna bagi pendidik, Lembaga Pendidikan, dan juga stakeholder lainnya. Tidak hanya itu, hal tersebut juga membantu dalam perbaikan kurikulum, metode pembelajaran, dan juga strategi evaluasi guna meningkatkan efektifitas pembelajaran PAI sendiri. Sehingga validitas memegang peranan yang penting dalam sistem evaluasi pendidikan guna memperoleh hasil yang akurat dan dapat dipertanggung jawabkan.

Validitas alat ukur dalam evaluasi setidaknya harus memenuhi dua aspek utama yaitu; validitas isi (*Content Validity*) untuk memastikan kesesuaian materi dengan kurikulum, dan Validitas Konstruksi (*Construct Validity*) untuk memastikan instrumen tersebut tepat dalam menggambarkan struktur berpikir atau sikap siswa. Apabila kedua aspek ini diabaikan, maka evaluasi pendidikan hanya akan menjadi formalitas administratif yang kehilangan esensi transformatifnya. Berdasarkan urgensi tersebut, makalah ini disusun untuk membedah lebih mendalam mengenai peran dan implementasi validitas dalam alat ukur evaluasi, guna mewujudkan sistem penilaian PAI yang kredibel dan objektif.

Selain validitas, reliabilitas juga merupakan aspek penting yang harus diperhatikan dalam penyusunan instrumen evaluasi. Reliabilitas berkaitan dengan tingkat konsistensi hasil pengukuran yang diperoleh ketika instrumen digunakan berulang kali pada kondisi yang relatif sama. Instrumen yang reliabel akan menghasilkan data yang stabil dan dapat dipercaya sehingga mampu memberikan gambaran yang akurat mengenai kemampuan peserta didik. Dalam konteks Pendidikan Agama Islam (PAI), reliabilitas menjadi penting karena hasil evaluasi sering dijadikan dasar dalam pengambilan keputusan pendidikan, seperti penentuan tingkat pencapaian kompetensi, perbaikan proses pembelajaran, maupun pengembangan kurikulum. Instrumen yang memiliki validitas tinggi tetapi tidak reliabel dapat menghasilkan data yang tidak konsisten sehingga mengurangi keakuratan hasil penilaian. Oleh karena itu, validitas dan reliabilitas merupakan dua komponen yang saling melengkapi dalam menjamin kualitas instrumen evaluasi pembelajaran.

Berdasarkan urgensi tersebut, kajian ini disusun untuk membedah lebih mendalam mengenai konsep validitas dan reliabilitas instrumen evaluasi pembelajaran, khususnya dalam Pendidikan Agama Islam. Pembahasan ini diharapkan dapat memberikan pemahaman yang komprehensif mengenai pentingnya kedua aspek tersebut dalam mewujudkan sistem penilaian yang objektif, akurat, konsisten, dan dapat dipertanggungjawabkan secara ilmiah.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan jenis penelitian kepustakaan (*library research*). Penelitian kepustakaan merupakan metode penelitian yang memanfaatkan berbagai sumber tertulis sebagai data utama untuk memperoleh pemahaman yang komprehensif mengenai suatu konsep atau fenomena. Fokus kajian dalam penelitian ini adalah konsep validitas dan reliabilitas instrumen evaluasi pembelajaran, khususnya dalam konteks Pendidikan Agama Islam (PAI).

⁷ La Zubair et al., "Strategi Inovatif Dalam Pengembangan Evaluasi Pembelajaran Pendidikan Agama Islam Untuk Meningkatkan Kualitas Pendidikan," *Jurnal Pendidikan Indonesia* 5, no. 11 (2024): 1217–27.

Sumber data yang digunakan terdiri atas sumber primer dan sumber sekunder. Sumber primer meliputi buku-buku evaluasi pendidikan, pengukuran dan penilaian pembelajaran, serta literatur yang secara khusus membahas validitas dan reliabilitas instrumen. Adapun sumber sekunder berupa artikel jurnal ilmiah, prosiding, dan dokumen akademik lain yang relevan dengan topik penelitian. Seluruh sumber dipilih berdasarkan tingkat relevansi, kredibilitas, dan keterkinian informasi yang disajikan.

Teknik pengumpulan data dilakukan melalui studi dokumentasi dengan cara mengidentifikasi, mengumpulkan, membaca, dan mengkaji berbagai literatur yang berkaitan dengan validitas dan reliabilitas instrumen evaluasi pembelajaran. Data yang diperoleh kemudian diklasifikasikan berdasarkan tema-tema pembahasan, seperti konsep validitas, jenis-jenis validitas, konsep reliabilitas, teknik pengujian reliabilitas, faktor-faktor yang memengaruhi reliabilitas, serta implikasinya dalam evaluasi pendidikan.

Analisis data dilakukan menggunakan teknik analisis isi (*content analysis*). Analisis dilakukan melalui tahapan reduksi data, penyajian data, interpretasi, dan penarikan kesimpulan. Pada tahap reduksi data, peneliti menyeleksi informasi yang relevan dengan fokus penelitian. Selanjutnya data disajikan secara sistematis berdasarkan kategori pembahasan, kemudian diinterpretasikan dengan menghubungkan berbagai pandangan para ahli untuk memperoleh pemahaman yang utuh mengenai validitas dan reliabilitas instrumen evaluasi. Untuk menjaga keabsahan data, peneliti menerapkan teknik triangulasi sumber dengan membandingkan berbagai referensi yang memiliki kesesuaian tema sehingga diperoleh informasi yang lebih objektif dan dapat dipertanggungjawabkan secara akademik.

Metode ini dipilih karena memungkinkan peneliti mengkaji secara mendalam konsep-konsep teoritis mengenai validitas dan reliabilitas instrumen evaluasi serta relevansinya dalam meningkatkan kualitas evaluasi pembelajaran Pendidikan Agama Islam.

HASIL DAN PEMBAHASAN

Definisi Konsep Validitas

Validitas berasal dari istilah Inggris, yaitu "validity," yang mengacu pada sejauh mana akurasi dan kecermatan suatu instrumen pengukur dalam menjalankan fungsinya⁸. Dalam konteks pengujian instrumen pengumpulan data, Azwar menjelaskan bahwa validitas mengindikasikan sejauh mana ketepatan dan kecermatan instrument pengukur (tes) dalam melaksanakan fungsinya sebagai alat ukur⁹. Dengan demikian, dapat dimaknai bahwa validitas mencerminkan sejauh mana sebuah alat ukur secara akurat mengukur aspek yang ingin diukur.

Validitas dalam evaluasi pembelajaran merujuk pada sejauh mana alat ukur atau instrumen yang digunakan mampu mengukur apa yang sebenarnya hendak diukur¹⁰. Dengan kata lain, validitas menjamin bahwa hasil evaluasi mencerminkan kemampuan, pengetahuan, atau kompetensi yang menjadi tujuan pembelajaran secara akurat dan tepat. Misalnya, jika tujuan evaluasi adalah mengukur pemahaman konsep tatacara shalat, maka soal atau tugas yang diberikan harus benar-benar berkaitan dengan konsep tersebut, bukan hanya menguji kemampuan menghafal bacaan shalat tanpa mengaitkan

⁸ Rosyidah Nur Ainy Sanusi and Furqanul Aziez, "Analisis Butir Soal Tes Objektif Dan Subjektif Untuk Keterampilan Membaca Pemahaman Pada Kelas Vii Smp n 3 Kalibagor," *Metafora: Jurnal Pembelajaran Bahasa Dan Sastra* 8, no. 1 (2021): 99–109.

⁹ Azwar S., *Tes Prestasi: Fungsi Dan Pengembangan Pengukuran Prestasi Belajar* (Yogyakarta: Pustaka Pelajar, 2011).

¹⁰ Nurul Muslikhah, "VALIDITAS, RELIABILITAS, OBJEKTIVITAS DAN AKUNTABILITAS DALAM EVALUASI PEMBELAJARAN," n.d.

pemahaman konsepnya. Validitas menjamin bahwa hubungan antara tujuan evaluasi dan isi yang diukur itu selaras dan tepat sasaran.

Validitas dalam evaluasi pembelajaran ditegaskan oleh Downing dan Haladyna yang menyatakan bahwa evaluasi yang valid berkontribusi pada peningkatan kualitas pembelajaran karena hasil evaluasi dapat dijadikan umpan balik yang tepat bagi guru dan peserta didik¹¹. Validitas yang terjamin memungkinkan guru untuk mengidentifikasi kekuatan dan kelemahan peserta didik secara akurat sehingga strategi pembelajaran dapat disesuaikan dengan kebutuhan nyata¹². Selain itu, validitas juga mendukung keadilan dalam proses evaluasi, memastikan bahwa semua peserta didik dinilai berdasarkan standar kompetensi yang sama tanpa adanya bias atau distorsi. Berdasarkan hal tersebut, validitas memiliki manfaat dalam evaluasi pembelajaran sebagai berikut:

1. Menghasilkan Data yang Akurat: Instrumen yang valid memastikan bahwa data yang diperoleh benar-benar mencerminkan pencapaian kompetensi siswa. Hal ini penting untuk pengambilan keputusan yang tepat mengenai perkembangan belajar siswa ataupun perbaikan kurikulum.
2. Mendukung Keadilan Penilaian: Dengan instrumen yang valid, setiap siswa dinilai berdasarkan kompetensi yang telah ditentukan, sehingga penilaian menjadi lebih adil dan tidak dipengaruhi oleh materi yang kurang relevan.
3. Dasar Perbaikan Proses Pembelajaran: Validitas membantu pendidik dan manajemen sekolah untuk mengidentifikasi area-area pengajaran yang perlu ditingkatkan. Data evaluasi yang valid menjadi dasar untuk merancang strategi pengajaran yang lebih efektif dan inovatif.

Dalam praktik evaluasi pembelajaran, memastikan validitas seringkali menghadirkan tantangan. Faktor-faktor seperti perbedaan interpretasi soal oleh siswa, keberagaman kemampuan, dan keterbatasan waktu dalam menyusun instrumen dapat mempengaruhi validitas. Oleh karena itu, penting bagi pendidik untuk secara rutin melakukan validasi instrumen, misalnya dengan uji coba (trial) dan review oleh ahli mata pelajaran guna memastikan kesesuaiannya dengan tujuan yang diinginkan.

Praktik penerapannya, validitas harus diuji dan dipastikan melalui berbagai teknik validasi, seperti validasi ahli, uji coba instrumen, dan analisis statistik¹³. Pengujian validitas ini harus menjadi bagian integral dari pengembangan instrumen evaluasi agar hasil yang diperoleh dapat dipertanggungjawabkan secara ilmiah. Selain itu, validitas harus dilihat sebagai proses dinamis yang memerlukan revisi dan penyesuaian berkelanjutan sesuai dengan perkembangan kurikulum dan konteks pembelajaran. Dengan demikian, penerapan validitas dalam evaluasi pembelajaran tidak hanya meningkatkan kualitas pengukuran, tetapi juga mendukung efektivitas proses pembelajaran secara keseluruhan.

Jenis-Jenis Validitas dalam Evaluasi

Secara umum, dalam mengevaluasi validitas alat ukur, yang berarti sejauh mana alat tersebut mencapai tujuan pengukuran, melibatkan tiga aspek, yakni: a) isi yang diukur, b)

¹¹ Thomas Downing, Ronald; Haladyna, *Assessment in Education: Principles and Practices* (New York: Pearson, 2021).

¹² Maulidani Ulfah, Darmansyah Darmansyah, and Rehani Rehani, "Instrumen Pengujian Produk Pembelajaran (Pengujian Validitas, Praktikalitas, Efektivitas)," *At-Tarbiyah: Jurnal Penelitian Dan Pendidikan Agama Islam* 3, no. 1 (2025): 43–51.

¹³ Upchurch Smith, James; Laura, *Techniques in Educational Assessment and Validation* (Boston: Academic Press, 2022).

rekaan teoritis (konstruk) dari atribut yang diukur, dan c) kriteria yang digunakan oleh alat ukur¹⁴.

a. Validitas Isi (*Content Validity*):

Validitas isi dalam dunia pendidikan memastikan sejauh mana butir-butir soal dalam suatu instrumen evaluasi mampu mewakili keseluruhan kawasan isi (*content domain*) objek yang hendak diukur. Validitas isi memastikan bahwa tes yang diberikan kepada siswa benar-benar mencerminkan kurikulum atau materi ajar yang telah disampaikan. Oleh karena materi yang diajarkan tertera dalam kurikulum, maka validitas isi sering juga disebut validitas kurikuler¹⁵. Contohnya, dalam satu semester guru PAI 5 Bab yang meliputi bidang aqidah, akhlak, fiqih, sejarah, dan Al-Qur'an, maka ketika ujian, soal-soalnya harus mencakup 5 bidang tersebut dalam artian mewakili semua isi pelajaran. Jika saat ujian soalnya hanya tentang fiqih saja, maka tes tersebut tidak memiliki validitas isi yang baik karena bidang lainnya ditinggal.

Validitas isi merujuk pada seberapa tepat suatu alat pengukur hasil belajar dengan isi dari materi yang diukur¹⁶. Validitas isi juga sering disebut sebagai validitas kurikulum, yang berarti bahwa suatu alat pengukur dianggap valid jika sesuai dengan isi kurikulum yang menjadi dasar pengukuran¹⁷. Dengan demikian, Validitas isi erat kaitannya dengan materi yang akan diukur dalam tes. Tentu saja materi yang dimaksud adalah materi yang terdapat dalam kurikulum. Validitas isi mencerminkan sejauh mana butir-butir dalam tes mencerminkan materi yang disajikan dalam kurikulum.

Pendekatan rasional digunakan untuk mengevaluasi validitas isi, dengan membandingkan soal dengan kisi-kisi soal atau kurikulum yang telah dipelajari¹⁸. Sebuah tes dikatakan memiliki validitas isi jika butir-butir tes bersifat representatif terhadap isi materi dalam kurikulum tersebut. Cara yang praktis untuk melakukan analisis validitas isi adalah dengan melihat apakah butir-butir tes telah disusun sesuai dengan blue-print (kisi-kisi) yang sudah dirancang sebelumnya. Blue print menjadi acuan dalam menuangkan domain atau ranah dan indikator yang akan diukur dalam tes.

Validitas isi berfungsi untuk memastikan soal-soal tidak keluar dari kurikulum. Validitas isi juga menekankan aspek keadilan dalam penerapannya, karena peserta didik diuji berdasarkan apa yang telah diajarkan di kelas, sehingga nilai hasilnya benar-benar mencerminkan usaha belajar mereka. Tidak hanya itu, validitas isi memberikan umpan balik yang akurat kepada guru mengenai bagian materi mana yang sudah dikuasai atau belum dikuasai oleh peserta didik. Hal ini kemudian menjadi pegangan guru untuk perbaikan agar pembelajaran menjadi lebih efektif kedepannya. Validitas isi dalam Pendidikan Agama Islam memiliki kedudukan yang strategis karena menjaga kemurnian ajaran. Dalam hal ini memastikan soal-soal tidak keluar dari koridor normativ Al-Qur'an dan As-Sunnah.

Validitas isi dapat diusahakan tercapainya sejak penyusunan dengan memerinci materi kurikulum atau materi buku pelajaran. Untuk menghasilkan instrument atau alat ukur yang memiliki validitas isi yang tinggi, berikut langkah-langkah yang bisa

¹⁴ Endro Suseno, S Kom, and MPd M Pd Dr Purwo Susongko, *Mengukur Validitas Tes* (Pemeral edukreatif, 2021).

¹⁵ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan Edisi 3* (Bumi aksara, 2021).

¹⁶ Anas Sudijono, "Pengantar Evaluasi Pendidikan," 2013.

¹⁷ Surapranata Sumarna, "Analisis, Validitas, Reliabilitas, Dan Interpretasi Hasil Tes Implementasi Kurikulum 2004," *Bandung: Remaja Rosdakarya*, 2005.

¹⁸ M. Chabib. Thoha, *Teknik Evaluasi Pendidikan* (Jakarta: PT. RajaGrafindo Persada-Rajawali Pers, 2003).

dilakukan, yaitu: (1) Analisis Kurikulum dan Silabus, (2) Penyusunan Kisi-kisi (*Blueprint*), (3) Verifikasi Kriteria Konsep, (4) Validasi Pakar (*Professional Judgment*).

b. Validitas Konstruk (*Construct Validity*):

Dalam evaluasi pembelajaran, validitas konstruk mempertanyakan apakah butir-butir soal dalam tes itu telah sesuai dengan tingkatan kompetensi atau ranah yang ada sesuai yang dituntut dalam kurikulum. Menilai apakah instrumen tersebut benar-benar mengukur konstruk teoritis yang dimaksud, seperti kreativitas, pemecahan masalah, atau kemampuan berpikir kritis (lebih kepada mengukur tingkat pemahaman). Validitas konstruk memerlukan dasar teoritis dan bukti empiris bahwa instrumen yang digunakan memang menangkap variabel yang ingin diukur. Contohnya: guru PAI ingin mengukur apakah peserta didik paham konsep tauhid. Akan tetapi, soal yang dibuat adalah "sebutkan rukun-rukun iman!", maka ini hanya sekedar mengukur hafalan peserta didik, bukan pemahamannya. Dengan semikian, soal yang dibuat oleh guru tersebut dapat dikatakan lemah secara validitas konstruk karena tidak berhasil menggambarkan pemahaman peserta didik terkait konsep tauhid.

Analisis validitas konstruk suatu tes dapat dilakukan dengan melakukan pencocokan antara kemampuan berpikir yang terkandung dalam butir-butir soal dengan dengan kemampuan berpikir yang tercantum dalam setiap rumusan indikator yang akan diukur. Sebagaimana dalam validitas isi, cara analisis dapat pula dilakukan dengan melakukan diskusi dengan orang yang ahli di bidang yang ber-sangkutan. Dengan kata lain, uji validitas konstruk dilakukan dengan cara expert judgement. Saat ini, uji validitas konstruk juga sering dilakukan lewat program komputer, yaitu dengan menggunakan analisis faktor. Jika cara ini yang dipakai, uji validitas tersebut harus berdasarkan data-data empirik. Hal ini berarti alat tes tersebut harus diujicobakan terlebih dahulu, dan data-data hasil uji coba itulah yang kemudian dianalisis dengan komputer.

Dalam Pendidikan Agama Islam, validitas konstruk menjadi sangat krusial karena objek ukur utama sering kali berupa nilai-nilai spiritual dan mental, seperti sikap religiusitas atau pemahaman konsep tauhid. Olehkarena itu, menjadi hal penting menyusun definisi operasional untuk mengubah konsep abstrak dalam PAI menjadi indikator yang dapat diamati dan diukur secara empiris. Definisi operasional ini berfungsi mengubah konsep abstrak teologis dalam PAI menjadi indikator-indikator perilaku yang dapat diamati dan diukur secara empirik. Hal ini termasuk salah satu langkah kunci dalam mencapai validitas konstruksi. Sebagai contoh misalnya tentang konsep akhlakul karimah, dibuat devinisi operasional menjadi indikator jujur dalam ujian, santun dalam berbicara, dan disiplin waktu. Penyusunan devinisi operasional harus tepat dengan melihat pada kurikulum beserta turunannya baik tujuan pebelajaran (TP) ataupun capaian pembelajaran (CP). Definisi operasional yang tepat mencegah terjadinya ambiguitas dalam penilaian dan memastikan bahwa setiap butir soal memiliki sandaran konsep yang jelas.

Penyusunan instrument yang memiliki validitas konstruksi yang baik menuntut adanya pemahaman terhadap struktur teoritis kognitif peserta didik. Struktur teoritis kognitif peserta didik ini, merujuk pada taksonomi Bloom pada domain kognitif/pengetahuan yang berkaitan dengan ingatan, berpikir, dan proses-proses penalaran. Sebagaimana disebutkan dalam taksonomi Bloom bahwa ada 6 tahapan, yaitu: pengetahuan, pemahaman, aplikasik, analisis, sintesis, dan evaluasi. Dari hal itu, guru harus mampu menyusun butir soal yang membedakan secara tegas antara tingkat pengetahuan/hafalan (*Lower Order Thinking Skills*) dengan tingkat analisis (*Higher Order Thinking Skills*) sesuai dengan taksonomi pendidikan seperti Taksonomi Bloom ini.

Validitas konstruk pada level satu (C1) yaitu pengetahuan, dapat dibuat soal/pertanyaan yang mengarah pada ingatan atau hafalan terkait teori yang dipelajari. Berbeda dengan level empat (C4) yaitu analisis, mengharuskan soal yang mamou memicu siswa untuk menguraikan dan mengorganisasi teori yang dipelajari, bukan sekedar menggali memori/ingatan (C1). Guru harus mampu membuat soal berdasarkan tingkatan kognitif peserta didik. Tidak hanya berhenti di situ, instrumen juga harus dirancang secara bertahap sesuai dengan perkembangan psikologis peserta didik agar konsep abstrak dapat disesuaikan dengan kapasitas kognitif mereka.

c. Validitas Kriteria (*Criterion-related Validity*):

Mengukur hubungan antara hasil evaluasi dengan suatu standar atau kriteria eksternal yang telah terbukti. Misalnya, guru PAI ingin mengukur kemampuan membaca Al-Qur'an peserta didik. Guru tersebut membuat instrument tes tulis tentang hukun-hukum tajwid, kemudian untuk menguji validitas kriteria, maka guru tersebut membandingkan nilai tes tersebut dengan praktik mengaji di depan guru tahfidz (sebagai kriteria standar). Jika peserta didik mendapatkan hasil nilai teorinya bagus dan ternyata praktiknya juga fasih, maka tes tertulis tersebut valid secara kriteria. Akan tetapi jika teorinya bagus dan praktiknya tidak, maka tes tulis tersebut lemah validitas kriterianya karena tidak sesuai dengan kenyataan (kriteria) di lapangan.

Jenis Validitas	Fokus Utama	Cara Mengecek
Validitas Isi	Materi Pelajaran	Menyesuaikan soal dengan kisi-kisi dan kurikulum.
Validitas Konstruk	Konsep/Cara Berpikir	Memastikan soal mengukur pemahaman, bukan cuma hafalan.
Validitas Kriteria	Perbandingan Luar	Membandingkan skor tes dengan standar lain atau prestasi nyata.

Konsep Reliabilitas Tes Hasil Belajar

Reliabilitas berasal dari kata "reliable", yang berarti "dapat dipercaya". Reliabilitas didefinisikan secara teknis sebagai proporsi keragaman skor tes yang disebabkan oleh keragaman kemampuan peserta, bukan kesalahan pengukuran acak.¹⁹ Dalam teori ujian klasik (CTT), skor perolehan sama dengan skor murni ditambah skor kesalahan. Kesalahan pengukura, dalam teori pengukuran klasik adalah perbedaan antara skor siswa dengan kemampuan sebenarnya mereka. Keadaan fisik peserta saat ujian, lingkungan ujian, ketidak jelasan instrumen, dan subjektivitas penskoran adalah beberapa dari banyak penyebab kesalahan ini.

¹⁹ Sumarna Surapranata, *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes* (Bandung: Remaja Rosdakarya, 2006), hlm. 89.

Instrumen tes yang lebih reliabel memiliki tingkat kesalahan pengukuran yang lebih rendah.²⁰ Alat dianggap reliabel jika koefisien reliabilitas (r_{11}) lebih dari 0,70, yang berarti bahwa setidaknya 70% dari variasi skor menunjukkan kemampuan nyata peserta daripada kesalahan acak.²¹ Secara umum, koefisien reliabilitas terbagi menjadi beberapa kategori: koefisien di bawah 0,50 menunjukkan bahwa itu kurang reliabel, koefisien antara 0,50–0,70 menunjukkan tingkat reliabilitas sedang, dan koefisien di atas 0,70 menunjukkan bahwa itu sangat reliabel. Semakin mendekati angka 1,00, instrumen semakin konsisten. Kesalahan pengukuran, atau kesalahan pengukuran, dalam teori pengukuran klasik adalah perbedaan antara skor siswa dengan kemampuan sebenarnya mereka. Keadaan fisik peserta saat ujian, lingkungan ujian, ketidakjelasan instrumen, dan subjektivitas penskoran adalah beberapa dari banyak penyebab kesalahan ini. Instrumen tes yang lebih reliabel memiliki tingkat kesalahan pengukuran yang lebih rendah.

Reliabilitas bergantung pada karakteristik populasi. Tes yang sama dapat memiliki koefisien reliabilitas yang berbeda untuk kelompok populasi yang berbeda (Supriadi, 2006). Sebagai contoh, apabila seorang siswa mengikuti ujian matematika dua kali dan memperoleh hasil yang hampir sama, maka ujian tersebut dianggap memiliki reliabilitas yang baik. Sebaliknya, jika hasil ujian berubah secara signifikan tanpa mengubah kemampuan siswa, instrumen tersebut dianggap kurang reliabel. Oleh karena itu, ketika koefisien reliabilitas dihitung, konteks populasi harus dipertimbangkan. Dalam evaluasi pendidikan, reliabilitas menentukan tingkat kepercayaan terhadap hasil pengukuran. Instrumen yang reliabel dapat menghasilkan skor yang konsisten, yang memungkinkan pengambilan keputusan pendidikan yang lebih adil. Sebaliknya, instrumen yang tidak reliabel dapat menghasilkan data yang tidak stabil, yang dapat menyebabkan kesalahan dalam penilaian kemampuan siswa. Oleh karena itu, reliabilitas adalah salah satu persyaratan penting yang harus dipenuhi saat membuat alat evaluasi pembelajaran. Tidak hanya penting untuk memahami definisi reliabilitas, tetapi juga penting untuk memahami hubungan antara reliabilitas dan validitas dalam evaluasi pendidikan.

Hubungan antara validitas dan reliabilitas adalah hierarkis: sebuah tes yang valid pasti reliabel, tetapi belum tentu valid. Dengan kata lain, reliabilitas meskipun tidak cukup untuk validitas.²² Ini berarti bahwa untuk membuat alat evaluasi yang berkualitas, pengujian validitas dan reliabilitas harus selalu dilakukan bersamaan.

Berbagai pendapat menyatakan bahwa reliabilitas tidak hanya berkaitan dengan konsistensi hasil tes, tetapi juga mencerminkan kualitas instrumen evaluasi secara keseluruhan. Instrumen evaluasi yang reliabel akan membantu guru memberikan gambaran yang lebih akurat tentang kemampuan siswa mereka. Oleh karena itu, pengujian reliabilitas seharusnya menjadi langkah penting dalam proses pembuatan instrumen evaluasi pembelajaran. Pengujian reliabilitas seringkali tidak dilakukan secara efektif dalam praktik pembelajaran di sekolah karena banyak guru masih menggunakan alat evaluasi tanpa melakukan analisis reliabilitasnya terlebih dahulu. Padahal, alat yang tidak reliabel dapat memengaruhi keakuratan hasil penilaian dan membuat keputusan yang tidak akurat tentang kemampuan siswa mereka.

Teknik Pengujian Reliabilitas

²⁰ Anas Sudijono, *Pengantar Evaluasi Pendidikan* (Ed. 1, Cet. 14; Jakarta: Rajawali Pers, 2015), hlm. 207–213.

²¹ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2018), hlm. 87.

²² Wawan Arbeni dkk., "Test Reliability Analysis in Educational Evaluation: A Quantitative Approach to Consistency and Validity," *Holistic Science*, Vol. 5 No. 1 (2025), hlm. 60–63.

Terdapat beragam teknik pengujian reliabilitas yang dapat dipilih sesuai dengan karakteristik instrumen dan konteks penggunaannya. Berikut ini disajikan analisis masing-masing teknik secara kritis.

1. Test-Retest (Stabilitas)

Metode tes-retest melibatkan pemberian tes yang sama kepada kelompok peserta yang sama pada dua titik waktu yang berbeda, dan kemudian menghitung hubungan antara kedua perangkat skor. Skor yang stabil dari waktu ke waktu ditunjukkan oleh korelasi yang tinggi.²³ Metode ini dapat digunakan untuk mengukur reliabilitas stabilitas, yang berarti konsistensi hasil pengukuran lintas waktu.

Kelebihan: mudah dipahami dan diimplementasikan; tidak memerlukan dua versi tes. **Kelemahan:** rentan terhadap efek carry-over (peserta mengingat jawaban dari pengujian pertama); hasil dapat dipengaruhi perubahan kondisi responden (kematangan, kelelahan, kejadian di antara dua tes). Interval waktu yang terlalu pendek memiliki efek yang lebih besar pada memori, sedangkan interval waktu yang terlalu panjang meningkatkan kemungkinan mengubah kemampuan aktual.²⁴ Teknik test-retest lebih tepat untuk alat yang mengukur kemampuan yang relatif stabil dalam jangka waktu tertentu, seperti tes bakat atau kemampuan dasar. Namun, metode ini tidak cocok untuk kemampuan yang mudah berubah karena proses pembelajaran intensif.

2. Split-Half (Reliabilitas Belah Dua)

Reliabilitas instrumen dapat diukur melalui konsistensi internalnya dengan menggunakan metode split-half selain metode stabilitas melalui test-retest. Teknik pembagian separuh membagi tes (seperti butir ganjil-genap) menjadi dua bagian yang setara, dan kemudian menghitung hubungan antara kedua bagian tersebut. Untuk mengoreksi pemendekan tes, formula Spearman-Brown, Flanagan, atau Rulon digunakan untuk menghitung koefisien reliabilitas akhir.²⁵

Kelebihan: cukup praktis karena hanya perlu melakukan satu tes; tidak ada efek carry-over. **Kelemahan:** asumsi ekuivalensi kedua belahan seringkali tidak terpenuhi dalam praktik; nilai koefisien yang dihasilkan dapat dipengaruhi oleh cara pembagian tes, seperti ganjil-genap atau separuh awal-akhir. Metode ini lebih sesuai untuk tes homogen. Untuk instrumen dengan jumlah item yang cukup besar dan homogen sehingga pembagian item menjadi dua bagian tetap dapat menghasilkan distribusi soal yang seimbang, teknik split half lebih cocok digunakan.

3. Cronbach Alpha (Konsistensi Internal)

Cronbach Alpha, yang diciptakan sebagai hasil dari pengembangan metode reliabilitas, dianggap lebih fleksibel daripada metode split-half karena mampu mengukur konsistensi internal secara lebih menyeluruh. Cronbach Alpha (α) merupakan generalisasi dari teknik *split-half* untuk semua kemungkinan pembagian tes. Rumus dasarnya adalah:

$$r_{11} = [n/(n-1)] \times [1 - (\sum S_i^2 / S^2)]$$

²³ Anas Sudijono, *Pengantar Evaluasi Pendidikan* (Ed. 1, Cet. 14; Jakarta: Rajawali Pers, 2015), hlm. 207–213.

²⁴ Muhammad Khumaedi, "Reliabilitas Instrumen Penelitian Pendidikan," *Jurnal Pendidikan Teknik Mesin*, Vol. 12 No. 1 (2012), hlm. 3–4.

²⁵ Ida Ayu Gde Yadnyawati, *Evaluasi Pembelajaran* (Denpasar: Unhi Press, 2019), hlm. 124–126.

di mana n adalah jumlah butir, $\sum S_i^2$ adalah jumlah varians butir, dan S^2 adalah jumlah varians keseluruhan. . Karena kemampuan untuk mengukur konsistensi internal secara menyeluruh, teknik ini paling banyak digunakan dalam penelitian pendidikan dan psikologi (terutama untuk instrumen dengan butir berskala politomi (tes uraian, skala Likert). Sebagai contoh, instrumen tes uraian dengan lima butir diuji pada lima peserta menghasilkan koefisien $\alpha = 0,94$, yang menunjukkan bahwa instrumen tersebut sangat reliabel.²⁶

Kelebihan: luas, tidak memerlukan belahan tes, dan berlaku untuk berbagai jenis instrumen. **Kelemahan:** kurang efektif apabila item tes tidak homogen; menganggap bahwa semua item mengukur konstruk yang sama (unidimensional); sensitif terhadap jumlah item (lebih banyak item cenderung menghasilkan alpha yang lebih tinggi meskipun konsistensi tidak meningkat secara signifikan).²⁷ Karena dapat diterapkan pada berbagai jenis instrumen, seperti skala Likert, angket, dan tes uraian yang memiliki banyak item pengukuran, Cronbach Alpha sangat cocok untuk digunakan dalam penelitian pendidikan modern.

4. Kuder-Richardson 20 (KR-20) dan KR-21

Alat dengan skor dikotomi, KR-20 dan KR-21, menggunakan metode reliabilitas tambahan selain Cronbach Alpha. Tes dikotomi (jawaban benar-salah) menggunakan kasus Cronbach Alpha KR-20 dan KR-21. Sementara formula KR-20 menggunakan proporsi jawaban benar (p) dan salah ($q = 1 - p$) untuk setiap butir, formula KR-21 mengambil asumsi bahwa semua item memiliki tingkat kesulitan yang sama dan menggunakan rerata skor total (M) sebagai pengganti perhitungan per butir. Akibatnya, formula KR-21 lebih cepat dan lebih mudah dihitung daripada KR-20, tetapi menghasilkan estimasi reliabilitas yang lebih konservatif.²⁸

Kelebihan KR-20: lebih akurat dari KR-21 dan cocok untuk tes dikotomi objektif. Kelemahan: Perhitungan proporsi per butir yang lebih kompleks diperlukan. Kelebihan KR-21: sangat praktis dan efektif. Kelemahan: berpotensi meremehkan koefisien reliabilitas sebenarnya; hanya berlaku apabila tingkat kesulitan semua item relatif seragam (Ropii & Fahrurrozi, 2017). KR-20 dan KR-21 biasanya digunakan untuk tes objektif pilihan ganda atau benar-salah. Ini karena kedua metode ini dibuat untuk item dikotomi.

Tabel 1. Analisis Komparatif Teknik Pengujian Reliabilitas Tes Hasil Belajar

Teknik	Jenis Tes	Prinsip	Kelebihan Utama	Keterbatasan Utama	Kondisi Optimal
Test-Retest	Semua jenis	Stabilitas lintas waktu	Mudah dipahami & diterapkan	Efek carry-over; perubahan kondisi responden	Tes kemampuan stabil (bukan pengetahuan faktual)

²⁶ Anas Sudijono, *Pengantar Evaluasi Pendidikan* (Ed. 1, Cet. 14; Jakarta: Rajawali Pers, 2015), hlm. 210–213.

²⁷ Putu Gede Subhaktiyasa, "Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif," *Journal of Education Research*, Vol. 5 No. 4 (2024), h. 5602–5603.

²⁸ Nyoman Doni Pramana dkk., *Evaluasi Pendidikan* (Surabaya: Beta Offset, 2014), hlm. 100–102.

Split-Half	Dikotomi & politomi homogen	Konsistensi dua belahan	Satu kali tes, tanpa efek carry-over	Cara belah memengaruhi koefisien	Tes homogen dengan jumlah butir genap
Cronbach Alpha	Politomi (uraian, Likert)	Konsistensi internal komprehensif	Komprehensif; tidak perlu belahan tes	Kurang efektif untuk tes heterogen	Tes uraian & skala sikap
KR-20	Dikotomi (objektif)	Proporsi benar-salah per butir	Akurat untuk tes dikotomi	Perhitungan per butir lebih rumit	Tes pilihan ganda dengan variasi kesukaran
KR-21	Dikotomi (objektif)	Rerata skor total	Sangat praktis & efisien	Asumsi kesukaran seragam; estimasi konservatif	Tes objektif dengan kesukaran seragam

Sesuai dengan tujuan pengukuran dan jenis instrumen yang digunakan, masing-masing metode memiliki fitur yang berbeda dari berbagai teknik tersebut. Berbeda dengan Cronbach Alpha dan split half, test-retest lebih berkonsentrasi pada kestabilan skor dari waktu ke waktu. Karena fleksibel dan dapat diterapkan pada berbagai instrumen, Cronbach Alpha menjadi teknik yang paling banyak digunakan dalam penelitian pendidikan saat ini. Meskipun demikian, KR-20 dan KR-21 lebih cocok untuk ujian objektif dengan skor dikotomi. Oleh karena itu, untuk membuat koefisien reliabilitas yang diperoleh lebih akurat, teknik reliabilitas harus disesuaikan dengan fitur instrumen, tujuan evaluasi, dan jenis data yang dihasilkan.

Dalam praktik evaluasi pembelajaran, guru dan peneliti harus memahami bahwa tidak semua teknik reliabilitas dapat digunakan pada semua jenis instrumen. Pemilihan teknik yang salah dapat menghasilkan estimasi reliabilitas yang kurang akurat. Oleh karena itu, guru harus menyesuaikan metode pengujian reliabilitas dengan alat yang digunakan, apakah itu angket pembelajaran, tes objektif, atau tes uraian.

Menurut berbagai metode yang telah dibahas, reliabilitas tidak hanya ditentukan oleh koefisien statistik yang tinggi tetapi juga oleh kesesuaian teknik pengujian dengan fitur instrumen evaluasi. Semakin tepat teknik yang digunakan, semakin akurat hasil pengukuran. Oleh karena itu, memahami karakteristik masing-masing teknik reliabilitas sangat penting bagi guru dan peneliti pendidikan ketika mereka membuat alat evaluasi yang berkualitas.

Meskipun berbagai metode reliabilitas telah berkembang dan digunakan secara luas dalam evaluasi pendidikan, masih ada kecenderungan bahwa mereka digunakan secara tidak tepat dalam praktiknya. Banyak guru dan peneliti hanya memilih metode penghitungan yang paling sederhana tanpa mempertimbangkan fitur instrumen yang digunakan. Akibatnya, koefisien reliabilitas yang diperoleh seringkali tidak sepenuhnya mencerminkan kualitas instrumen evaluasi yang sebenarnya. Selain itu, penggunaan alat

statistik kontemporer kadang-kadang membuat proses pengujian reliabilitas hanya administratif tanpa pengetahuan konseptual yang cukup. Padahal, memilih metode reliabilitas yang tepat sangat penting untuk memastikan bahwa hasil evaluasi pembelajaran berkualitas tinggi.

Faktor yang Memengaruhi Reliabilitas

Reliabilitas bukanlah karakteristik instrumen yang bersifat mutlak dan tetap. Berbagai faktor yang berkaitan dengan kualitas instrumen dan karakteristik siswa dipengaruhi oleh koefisien reliabilitas yang tinggi dan rendah.²⁹ Oleh karena itu, penting untuk memahami faktor-faktor yang memengaruhi reliabilitas agar proses evaluasi pembelajaran dapat menghasilkan data yang lebih akurat dan konsisten. Menurut Gronlund, setidaknya terdapat empat faktor utama yang secara signifikan memengaruhi koefisien reliabilitas.³⁰

1. Panjang Tes (Jumlah Butir)

Terdapat hubungan positif antara jumlah butir soal dan koefisien reliabilitas. Semakin banyak butir soal, jumlah sampel perilaku yang diukur sebanding dengan jumlah butir soal yang diberikan. Akibatnya, faktor tebakan, atau asumsi, semakin kecil dan hasil tes semakin konsisten. Namun, penambahan bagian hanya dapat dilakukan jika bagian-bagian tersebut memiliki kualitas yang sama dan merupakan bagian dari struktur yang sama. Sebagai contoh, tes dengan empat puluh butir soal biasanya lebih reliabel daripada tes dengan sepuluh butir soal karena cakupan materi yang diukur lebih luas dan variasi jawaban yang diberikan oleh peserta.

2. Tingkat Kesukaran (Difficulty Index)

Tes yang terlalu mudah atau terlalu sulit cenderung menghasilkan skor yang sempit, yang berarti mereka memiliki variasi yang rendah, sehingga mereka tidak memiliki koefisien reliabilitas yang baik. Tes yang menghasilkan distribusi skor yang mendekati kurva normal (di mana indeks kesukaran antara 0,30 dan 0,70) dianggap sebagai tingkat kesukaran ideal karena distribusi ini memaksimalkan variasi skor, yang merupakan prasyarat reliabilitas tinggi. Sebaliknya, jika soal itu terlalu sederhana, hampir semua peserta akan mendapatkan skor tinggi. Akibatnya, variasi skor akan semakin kecil dan instrumen akan menjadi kurang reliabel.

3. Objektivitas Penskoran

Penskoran subjektif cenderung membuat skor tes kurang konsisten baik di antara penskoran maupun di antara penskoran. Dalam dimensi penskoran, instrumen subjektif (uraian) kurang reliabel daripada instrumen objektif (pilihan ganda). Rubrik penilaian yang baik dapat meningkatkan kredibilitas dan objektivitas tes uraian.³¹ Misalnya, ketika tes uraian tidak memiliki rubrik yang jelas, dua penilai dapat memberikan skor yang berbeda untuk jawaban yang sama, menurunkan tingkat konsistensi hasil.

4. Homogenitas dan Sebaran Skor

Koneksi internal yang lebih tinggi cenderung dihasilkan oleh instrumen dengan komponen yang mengukur struktur yang homogen. Sebaliknya, semakin

²⁹ Adhi Setiyawan, "Faktor-Faktor yang Mempengaruhi Reliabilitas Tes," *Jurnal An Nûr*, Vol. VI No. 2 (2014), hlm. 345–353.

³⁰ Muhammad Ropii dan Muh. Fahrurrozi, *Evaluasi Hasil Belajar* (NTB: Universitas Hamzanwadi Press, 2017), hlm. 137–138.

³¹ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2018), hlm. 109–111.

besar variasi kemampuan peserta dalam kelompok, atau sebaran skor yang lebih besar, semakin besar kemungkinan korelasi antar-pengujian, yang berarti koefisien reliabilitas yang lebih tinggi. Ini menunjukkan bahwa alat yang sama dapat menghasilkan koefisien reliabilitas yang berbeda untuk kelompok heterogenitas yang berbeda.³² Oleh karena itu, alat yang mengukur satu konstruk secara konsisten cenderung memiliki reliabilitas yang lebih tinggi daripada alat yang mengukur beberapa konstruk sekaligus.

Tabel 2. Faktor-Faktor yang Memengaruhi Reliabilitas dan Rekomendasi Praktis

Faktor	Pengaruh terhadap Reliabilitas	Kondisi Ideal	Rekomendasi Praktis
Panjang Tes	Positif: lebih banyak butir → reliabilitas lebih tinggi	Jumlah butir proporsional dengan cakupan konstruk	Tambah butir berkualitas; uji coba instrumen sebelum digunakan
Tingkat Kesukaran	Sedang paling baik; terlalu mudah/sulit menurunkan reliabilitas	Indeks kesukaran 0,30–0,70	Lakukan analisis butir; eliminasi butir terlalu mudah/sulit
Objektivitas Penskoran	Tinggi: tes objektif lebih reliabel dalam penskoran	Pedoman/rubrik penilaian yang jelas dan terstandar	Gunakan rubrik; latih penilai; uji inter-rater reliability
Homogenitas Soal	Tinggi: butir yang mengukur konstruk sama menghasilkan reliabilitas lebih tinggi	Butir mengukur satu konstruk dominan (unidimensional)	Analisis item; pastikan seluruh butir relevan dengan indikator
Sebaran Skor	Semakin besar variasi skor, potensi reliabilitas lebih tinggi	Distribusi skor mendekati normal	Gunakan kelompok heterogen; hindari sampel yang terlalu homogen

Dalam praktik pembelajaran, guru seringkali lebih fokus pada jumlah soal daripada kualitas item yang disusun. Namun, reliabilitas tidak hanya ditentukan oleh banyaknya soal; konsistensi penskoran dan kualitas konstruksi item juga merupakan faktor penting dalam evaluasi pembelajaran di sekolah.

Reliabilitas instrumen tidak hanya dipengaruhi oleh metode perhitungan statistik, tetapi juga oleh kualitas penyusunan instrumen evaluasi secara keseluruhan. Instrumen yang disusun dengan mempertimbangkan jumlah item, tingkat kesukaran, objektivitas penskoran, dan homogenitas konstruk lebih cenderung menghasilkan reliabilitas yang baik. Oleh karena itu, agar hasil evaluasi lebih akurat dan dapat dipercaya, guru dan peneliti pendidikan harus mempertimbangkan elemen-elemen tersebut sejak tahap perencanaan instrumen.

Tabel 3. Faktor-Faktor yang Memengaruhi Reliabilitas Instrumen

³² Gito Supriadi, "Reliabilitas Tes Hasil Belajar dan Aplikasinya dalam Evaluasi Pembelajaran," *Himmah*, Vol. VII No. 18 (2006), hlm. 51–52.

Faktor	Pengaruh Terhadap Reliabilitas	Cara Meningkatkan
Panjang Tes	<p>Karena sampel perilaku yang diukur lebih luas dan faktor tebakan yang lebih kecil, reliabilitas item cenderung meningkat dengan jumlah item yang lebih besar.</p> <p>Sangat percaya pada reliabilitas</p>	<p>Jangan menambah item hanya untuk meningkatkan jumlah tanpa mempertimbangkan kualitasnya; sebaliknya, tambahkan item berkualitas yang sama.</p>
Tingkat Kesukaran	<p>Faktor-faktor yang menurunkan koefisien reliabilitas termasuk masalah yang terlalu mudah atau terlalu sulit untuk mempersempit sebaran skor.</p> <p>Reliabilitas rendah jika terlalu mudah atau sulit.</p>	<p>Gunakan soal dengan tingkat kesukaran sedang (indeks 0,30–0,70). Lakukan analisis butir dan hapus item yang terlalu sulit.</p>
Objektivitas	<p>Penskoran yang subjektif menurunkan reliabilitas, terutama pada tes uraian, karena menghasilkan skor yang tidak konsisten antarpenskoran.</p> <p>Subjektivitas mengurangi konsistensi.</p>	<p>Untuk memastikan keseragaman penilaian, gunakan rubrik penilaian yang jelas dan terstandar. Latih penilai dan uji ketepatan antar penilai.</p>
Homogenitas	<p>Konstruksi homogen menghasilkan koefisien reliabilitas yang lebih tinggi dan konsistensi internal yang lebih baik.</p>	<p>Fokuskan pada satu konstruk per instrumen dan analisis item untuk memastikan bahwa semua item sesuai dengan indikator yang sama.</p>

Tabel tersebut menunjukkan bahwa banyak faktor yang saling berhubungan mempengaruhi reliabilitas instrumen. Oleh karena itu, penyusunan instrumen evaluasi harus dilakukan dengan hati-hati agar hasil pengukuran yang diperoleh lebih konsisten dan dapat dipercaya.

Implikasi Reliabilitas dalam Evaluasi Pendidikan

Reliabilitas instrumen dalam evaluasi pendidikan berkaitan dengan aspek teknis pengukuran serta kualitas keputusan pendidikan yang dibuat. Reliabilitas instrumen memiliki dampak langsung dan signifikan terhadap kualitas pengambilan keputusan pendidikan. Data yang tidak dapat diandalkan dari instrumen dapat menyebabkan keputusan yang tidak adil tentang hal-hal seperti kelulusan, ranking, dan seberapa efektif program pembelajaran. Sebagai contoh, instrumen ujian yang tidak dapat diandalkan dapat menyebabkan peserta didik memperoleh hasil yang berbeda meskipun mereka memiliki tingkat kemampuan yang hampir sama. Kondisi ini dapat memengaruhi keputusan guru tentang apakah siswa diterima atau ditempatkan di kelas.

Zahra dkk. menyatakan bahwa alat penilaian yang tidak reliabel tidak dapat menjadi dasar yang sah untuk mengevaluasi pencapaian kompetensi siswa.³³ Sementara itu, Maulana menemukan bahwa alat penilaian dengan koefisien reliabilitas rendah menghasilkan klasifikasi kemampuan yang tidak stabil. Artinya, peserta yang sebenarnya memiliki kemampuan yang sama dapat diklasifikasikan dengan cara yang berbeda hanya karena ketidakkonsistenan instrumen.³⁴

Reliabilitas dilihat dalam kerangka evaluasi kontemporer sebagai proses metodologis yang mencakup perencanaan instrumen, uji coba, analisis detail, dan revisi berkelanjutan, bukan hanya kalkulasi statistik. Sangat penting bagi guru dan peneliti pendidikan untuk memasukkan pengujian kredibilitas ke dalam proses pembuatan instrumen evaluasi yang standar.³⁵ Ini terutama berlaku untuk instrumen evaluasi yang akan digunakan dalam pengambilan keputusan pendidikan yang sangat memengaruhi siswa. Berdasarkan berbagai pendapat, reliabilitas seharusnya dilihat sebagai indikator kualitas proses evaluasi pembelajaran secara keseluruhan, bukan hanya sebagai hasil perhitungan statistik.³⁶ Alat yang tepat akan membantu guru dan peneliti mendapatkan data yang lebih konsisten, yang akan memungkinkan pengambilan keputusan pendidikan yang lebih adil dan adil.

Dalam praktiknya, terdapat beberapa masalah yang sering menyebabkan rendahnya reliabilitas instrumen. Permasalahan tersebut antara lain penggunaan metode reliabilitas yang tidak sesuai dengan jenis instrumen, penggunaan sampel yang terlalu kecil atau tidak representatif, serta tidak dilakukannya uji coba instrumen sebelum digunakan secara resmi.³⁷ Berbagai kondisi tersebut dapat menyebabkan estimasi reliabilitas tidak mencerminkan kondisi instrumen yang sebenarnya.³⁸ Oleh karena itu, guru perlu melakukan uji coba instrumen serta analisis reliabilitas sebelum tes digunakan dalam evaluasi pembelajaran. Selain meningkatkan kualitas hasil penilaian, langkah ini juga dapat membantu guru memastikan bahwa instrumen yang digunakan benar-benar mampu mengukur kemampuan peserta didik secara konsisten. Dengan demikian, reliabilitas merupakan komponen penting dari evaluasi pendidikan karena berkaitan langsung dengan keakuratan hasil pengukuran dan kualitas pengambilan keputusan pendidikan. Semakin kuat instrumen, semakin percaya hasil evaluasi.

Pengujian kredibilitas masih belum menjadi prioritas utama dalam proses evaluasi pendidikan di dunia nyata. Banyak guru lebih fokus pada administrasi pembelajaran dan penyusunan soal daripada melakukan analisis kualitas instrumen. Akibatnya, alat evaluasi sering digunakan secara langsung tanpa melalui proses uji coba atau penghitungan reliabilitas. Kondisi tersebut dapat menyebabkan hasil evaluasi pembelajaran menjadi tidak konsisten dan tidak dapat secara objektif menggambarkan kemampuan siswa. Pada akhirnya, keputusan pendidikan yang dibuat menggunakan alat yang tidak dapat diandalkan dapat mengganggu proses penilaian siswa.

³³ Gelbita Zahra dkk., "Implementasi Penggunaan Instrumen Tes untuk Mengukur Objek Evaluasi Kognitif Pelajaran Bahasa Indonesia di Sekolah Dasar," *Jurnal Penelitian dan Evaluasi Pendidikan*, Vol. 6 No. 4 (2025), hlm. 960–961

³⁴ Andi Maulana, "Analisis Validitas dan Reliabilitas Instrumen Penilaian," *Jurnal Kualita Pendidikan*, Vol. 3 No. 3 (2022), hlm. 45.

³⁵ Muhammad Khumaedi, "Reliabilitas Instrumen Penelitian Pendidikan," *Jurnal Pendidikan Teknik Mesin*, Vol. 12 No. 1 (2012), hlm. 3–4.

³⁶ Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2018), hlm. 87.

³⁷ Afifah Aulia Zayrin dkk., "Analisis Instrumen Penelitian Pendidikan," *Jurnal QOSIM*, Vol. 3 No. 2 (2025), h. 782–783.

³⁸ Faiz Zahfa dkk., "Validity and Reliability of Educational Evaluation Tests," *Jurnal Riset Ilmu Pendidikan*, Vol. 5 No. 1 (2025), hlm. 12.

Berdasarkan pembahasan yang telah dilakukan, jelas bahwa reliabilitas merupakan komponen penting dari sistem evaluasi pendidikan secara keseluruhan, bukan sekadar masalah teknis statistik. Alat yang dapat diandalkan akan menghasilkan data yang lebih stabil dan objektif, yang akan membantu pengambilan keputusan yang lebih adil tentang pendidikan. Oleh karena itu, peningkatan kemampuan guru untuk membuat dan menganalisis alat evaluasi harus menjadi prioritas utama dalam pengembangan profesionalisme pendidikan.

KESIMPULAN DAN SARAN

Kesimpulan

Validitas dan reliabilitas merupakan dua komponen fundamental dalam pengembangan instrumen evaluasi pembelajaran yang menentukan kualitas hasil pengukuran. Validitas berkaitan dengan ketepatan instrumen dalam mengukur kompetensi atau aspek yang seharusnya diukur, sedangkan reliabilitas berhubungan dengan tingkat konsistensi hasil pengukuran ketika instrumen digunakan dalam kondisi yang relatif sama. Instrumen yang baik harus memenuhi kedua aspek tersebut agar hasil evaluasi dapat dipercaya dan digunakan sebagai dasar pengambilan keputusan pendidikan yang tepat.

Kajian ini menunjukkan bahwa validitas instrumen dapat ditinjau melalui validitas isi, validitas konstruk, dan validitas kriteria yang masing-masing memiliki peran penting dalam menjamin kesesuaian instrumen dengan tujuan pembelajaran. Sementara itu, reliabilitas dapat diuji melalui berbagai teknik, seperti test-retest, split-half, Cronbach Alpha, serta Kuder-Richardson (KR-20 dan KR-21), yang penggunaannya harus disesuaikan dengan karakteristik instrumen dan jenis data yang dihasilkan. Selain itu, reliabilitas dipengaruhi oleh berbagai faktor, antara lain jumlah butir soal, tingkat kesukaran, objektivitas penskoran, homogenitas konstruk, dan sebaran skor peserta didik.

Dalam konteks Pendidikan Agama Islam, validitas dan reliabilitas memiliki peranan yang sangat penting karena evaluasi tidak hanya mengukur aspek kognitif, tetapi juga aspek afektif dan nilai-nilai keagamaan. Oleh sebab itu, instrumen evaluasi yang valid dan reliabel akan menghasilkan data yang lebih akurat, objektif, serta mampu memberikan gambaran yang sebenarnya mengenai pencapaian kompetensi peserta didik. Dengan demikian, pengujian validitas dan reliabilitas perlu menjadi bagian yang tidak terpisahkan dalam proses penyusunan instrumen evaluasi pembelajaran.

Saran

Pendidik perlu meningkatkan kompetensi dalam menyusun dan menganalisis instrumen evaluasi agar mampu menghasilkan alat ukur yang valid dan reliabel. Setiap instrumen yang akan digunakan sebaiknya melalui proses validasi dan uji reliabilitas terlebih dahulu untuk memastikan kualitas hasil pengukuran. Selain itu, lembaga pendidikan perlu memberikan pelatihan dan pendampingan kepada guru terkait pengembangan instrumen evaluasi yang sesuai dengan prinsip-prinsip pengukuran pendidikan. Bagi peneliti selanjutnya, disarankan untuk melakukan penelitian empiris mengenai penerapan validitas dan reliabilitas instrumen evaluasi pada berbagai jenjang pendidikan serta mengembangkan model evaluasi yang lebih adaptif terhadap kebutuhan pembelajaran abad ke-21 dan karakteristik Pendidikan Agama Islam.

DAFTAR RUJUKAN

- Arbeni, W., Windiani, A., Sihotang, D.S.B., Anggraini, N., Wulandari, S., & Nugroho, A. (2025). Test Reliability Analysis in Educational Evaluation: A Quantitative Approach to Consistency and Validity. *Holistic Science*, 5(1), 59–64.
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

- Arikunto, Suharsimi. *Dasar-Dasar Evaluasi Pendidikan Edisi 3*. Bumi aksara, 2021.
- Biantoro, Oktio Frenki, and Asep Rahmatullah. "Internalisasi Nilai-Nilai Pendidikan Agama Islam Dalam Pembinaan Moral Siswa Di Sekolah." *Pelita: Jurnal Studi Islam Mahasiswa UII Dalwa* 2, no. 2 (2025): 225–41.
- Downing, Ronald; Haladyna, Thomas. *Assessment in Education: Principles and Practices*. New York: Pearson, 2021.
- Fernando, Derri Adi, Hartatiana Hartatiana, and Fajri Ismail. "Pentingnya Validitas Dan Reliabilitas Instrument Evaluasi Pada Pembelajaran Pendidikan Agama Islam." *Raudhah Proud To Be Professionals: Jurnal Tarbiyah Islamiyah* 8, no. 3 (2023): 1110–21.
- Fitrianti, Leni. "Penguatan Evaluasi Pembelajaran Bagi Guru Pemula." *Al-Ihda': Jurnal Pendidikan Dan Pemikiran* 18, no. 1 (2023): 838–49.
- Hendrawan, Wawan, Rudi Haryadi, and Cucu Atikah. "Instrumen Penilaian Pendidikan Agama Islam." *JTPM (Jurnal Teknologi Pendidikan Dan Pembelajaran): Edutech and Intructional Research Journal* 11, no. 2 (2024).
- Kamila, Aisyah, and Prasetyo Budi Widodo. "Validitas Konstruk Dalam Psikometri: Peran Analisis Faktor Eksploratori Dan Konfirmatori." *Jurnal Psikologi Pendidikan Dan Pengembangan SDM* 13, no. 1 (2025): 1–7.
- Khumaedi, M. (2012). Reliabilitas Instrumen Penelitian Pendidikan. *Jurnal Pendidikan Teknik Mesin*, 12(1), 1–7.
- Laila, Laila, Alawiyah Nabila, and Eka Widyanti. "Konsep Dasar Evaluasi Pembelajaran." *Jurnal Manajemen Dan Pendidikan Agama Islam* 2, no. 5 (2024): 252–62.
- Maulana, A. (2022). Analisis Validitas dan Reliabilitas Instrumen Penilaian. *Jurnal Kualita Pendidikan*, 3(3), 42–50.
- Muslikhah, Nurul. "VALIDITAS, RELIABILITAS, OBJEKTIVITAS DAN AKUNTABILITAS DALAM EVALUASI PEMBELAJARAN," n.d.
- Nafiati, Dewi Amaliah. "Revisi Taksonomi Bloom: Kognitif, Afektif, Dan Psikomotorik." *Humanika, Kajian Ilmiah Mata Kuliah Umum* 21, no. 2 (2021): 151–72.
- Ondeng, Syarifuddin, Ulfiani Rahman, and others. "Ragam Variabel Dan Definisi Operasional Penelitian Pendidikan: Implikasi Metodologis Pendidikan Islam Dan Pendidikan Matematika." *Jurnal Pendidikan, Penciptaan Seni Dan Budaya* 1, no. 2 (2025): 80–88.
- Pramana, N.D., dkk. (2014). *Evaluasi Pendidikan*. Surabaya: Beta Offset.
- Ropii, M., & Fahrurrozi, M. (2017). *Evaluasi Hasil Belajar*. NTB: Universitas Hamzanwadi Press.
- S., Azwar. *Tes Prestasi: Fungsi Dan Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar, 2011.
- Sanusi, Rosyidah Nur Ainy, and Furqanul Aziez. "Analisis Butir Soal Tes Objektif Dan Subjektif Untuk Keterampilan Membaca Pemahaman Pada Kelas Vii Smp n 3 Kalibagor." *Metafora: Jurnal Pembelajaran Bahasa Dan Sastra* 8, no. 1 (2021): 99–109.
- Saputra, Anri, and others. "Pengembangan Instrumen Evaluasi." *Ar-Raudah: Jurnal Pendidikan Dan Keagamaan* 2, no. 4 (2025): 1–14.
- Setiyawan, A. (2014). Faktor-Faktor yang Mempengaruhi Reliabilitas Tes. *Jurnal An Nûr*, VI(2), 345–353.

- Smith, James; Laura, Upchurch. *Techniques in Educational Assessment and Validation*. Boston: Academic Press, 2022.
- Subhaktiyasa, P.G. (2024). Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Journal of Education Research*, 5(4), 5598–5605.
- Sudijono, Anas. "Pengantar Evaluasi Pendidikan," 2013.
- Sudijono, A. (2015). *Pengantar Evaluasi Pendidikan* (Ed. 1, Cet. 14). Jakarta: Rajawali Pers.
- Sumarna, Surapranata. "Analisis, Validitas, Reliabilitas, Dan Interpretasi Hasil Tes Implementasi Kurikulum 2004." *Bandung: Remaja Rosdakarya*, 2005.
- Supriadi, G. (2006). Reliabilitas Tes Hasil Belajar dan Aplikasinya dalam Evaluasi Pembelajaran. *Himmah*, VII(18), 49–60.
- Surapranata, S. (2006). *Analisis, Validitas, Reliabilitas, dan Interpretasi Hasil Tes*. Bandung: Remaja Rosdakarya.
- Suseno, Endro, S Kom, and MPd M Pd Dr Purwo Susongko. *Mengukur Validitas Tes*. Pernal edukreatif, 2021.
- Thoha, M. Chabib. *Teknik Evaluasi Pendidikan*. Jakarta: PT. RajaGrafindo Persada-Rajawali Pers, 2003.
- Ulfah, Maulidani, Darmansyah Darmansyah, and Rehani Rehani. "Instrumen Pengujian Produk Pembelajaran (Pengujian Validitas, Praktikalitas, Efektivitas)." *At-Tarbiyah: Jurnal Penelitian Dan Pendidikan Agama Islam* 3, no. 1 (2025): 43–51.
- Widiana, I Wayan, I Ketut Gading, I Made Tegeh, and Putu Aditya Antara. *Validasi Penyusunan Instrumen Penelitian Pendidikan*. PT. RajaGrafindo Persada-Rajawali Pers, 2023.
- Yadnyawati, I.A.G. (2019). *Evaluasi Pembelajaran*. Denpasar: Unhi Press.
- Yusrizal; & Rahmawati. *Tes Hasil Belajar*. Banda Aceh: Percetakan Bandardi Lagugob, 2022.
- Zahfa, F., dkk. (2025). Validity and Reliability of Educational Evaluation Tests. *Jurnal Riset Ilmu Pendidikan*, 5(1), 8–15.
- Zahra, G., dkk. (2025). Implementasi Penggunaan Instrumen Tes untuk Mengukur Objek Evaluasi Kognitif Pelajaran Bahasa Indonesia di Sekolah Dasar. *Jurnal Penelitian dan Evaluasi Pendidikan*, 6(4), 955–965.
- Zayrin, A.A., dkk. (2025). Analisis Instrumen Penelitian Pendidikan. *Jurnal QOSIM*, 3(2), 778–786.
- Zubair, La, Dian Amirul Mu Mini, Zikri Adib Kurnia, and Abdul Bashith. "Strategi Inovatif Dalam Pengembangan Evaluasi Pembelajaran Pendidikan Agama Islam Untuk Meningkatkan Kualitas Pendidikan." *Jurnal Pendidikan Indonesia* 5, no. 11 (2024): 1217–27.