

RELIABILITAS ALAT UKUR, JENIS-JENIS DAN FAKTOR-FAKTOR YANG MEMPENGARUHINYA

Defrijon¹, Eka Matra², M. Rizki Lazuardi³, Yogi Kurniadi⁴, Lidus Yardi⁵,
Julhadi⁶

defrijon05@gmail.com¹, ekamatra253@gmail.com², rizkilazuardi@gmail.com³,
yogikurniadi763@gmail.com⁴, lidusyardi78@gmail.com⁵, julhadi15@gmail.com⁶

Universitas Muhammadiyah Sumatera Barat

ABSTRAK

Evaluasi pembelajaran dalam pendidikan Islam memiliki peran sentral dalam menentukan sejauh mana tujuan pembelajaran tercapai. Evaluasi yang baik membutuhkan alat ukur yang andal dan relevan. Reliabilitas alat ukur menjadi salah satu indikator utama dalam memastikan hasil evaluasi dapat dipercaya dan digunakan untuk pengambilan keputusan. Namun, banyak faktor yang memengaruhi reliabilitas ini, seperti kemampuan evaluator, kondisi peserta didik, serta lingkungan pembelajaran. Oleh karena itu, diperlukan kajian mendalam untuk memahami berbagai aspek yang memengaruhi kualitas evaluasi dalam pendidikan Islam.

Kata Kunci: Evaluasi Pembelajaran, Pendidikan Islam, Jenis Evaluasi.

ABSTRACT

Evaluation in Islamic education is a critical element to measure the effectiveness of the learning process. This article discusses the reliability of evaluation tools, types of evaluation, and the factors influencing them. The study employs a qualitative method with a literature review approach. The results indicate that the reliability of evaluation tools is influenced by instrument validity, consistent application, and relevance to educational objectives. Other factors include evaluator competency, environment, and learner conditions. This article provides recommendations to enhance the quality of evaluation in Islamic education.

Keywords: Learning Evaluation, Islamic Education, Types Of Evaluation.

PENDAHULUAN

Evaluasi pembelajaran merupakan salah satu elemen penting dalam sistem pendidikan Islam karena bertujuan untuk menilai sejauh mana tujuan pembelajaran telah tercapai. Dalam konteks pendidikan Islam, evaluasi tidak hanya berfungsi untuk mengukur aspek kognitif peserta didik tetapi juga mencakup dimensi afektif dan psikomotorik yang sesuai dengan nilai-nilai Islam. Evaluasi yang efektif membutuhkan alat ukur yang andal, valid, serta relevan dengan tujuan pembelajaran yang telah ditetapkan. Reliabilitas alat ukur menjadi salah satu indikator utama dalam memastikan bahwa hasil evaluasi dapat dipercaya dan memberikan informasi yang akurat untuk pengambilan keputusan. Reliabilitas mengacu pada konsistensi hasil evaluasi ketika alat ukur digunakan berulang kali dalam situasi yang serupa. Dengan demikian, reliabilitas yang tinggi memungkinkan pendidik untuk memperoleh hasil evaluasi yang dapat diandalkan sebagai dasar untuk merancang strategi pembelajaran berikutnya.

Namun, terdapat berbagai faktor yang dapat memengaruhi reliabilitas alat ukur dalam evaluasi pendidikan Islam. Pertama, kemampuan evaluator sangat menentukan hasil evaluasi. Evaluator yang kompeten akan mampu menyusun dan menggunakan alat ukur yang sesuai dengan kompetensi yang diharapkan. Kedua, kondisi peserta didik juga berperan penting. Keadaan emosional, fisik, dan motivasi peserta didik saat proses evaluasi dapat memengaruhi hasil yang diperoleh. Ketiga, lingkungan pembelajaran menjadi faktor eksternal yang tidak dapat diabaikan. Suasana yang mendukung akan

membantu peserta didik untuk fokus dan menunjukkan kemampuan terbaik mereka.

Oleh karena itu, kajian mendalam diperlukan untuk memahami berbagai aspek yang memengaruhi kualitas evaluasi dalam pendidikan Islam. Evaluasi yang bermutu tidak hanya berlandaskan pada pemilihan alat ukur yang tepat, tetapi juga mempertimbangkan konteks dan karakteristik peserta didik serta tujuan pendidikan itu sendiri. Dengan pendekatan yang holistik, evaluasi dalam pendidikan Islam dapat menjadi alat yang efektif untuk meningkatkan kualitas pembelajaran dan mencapai tujuan pendidikan Islam secara menyeluruh.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan metode studi pustaka (library research). Pendekatan ini dipilih karena bertujuan untuk menggali pemahaman mendalam tentang konsep, prinsip, dan praktik evaluasi pembelajaran dalam pendidikan Islam. Studi pustaka memungkinkan peneliti untuk menganalisis berbagai sumber literatur yang relevan, seperti jurnal ilmiah, buku referensi, dan dokumen resmi, sehingga dapat memperoleh gambaran yang komprehensif mengenai topik yang dikaji. Sebagaimana dinyatakan oleh Creswell pendekatan kualitatif sangat cocok digunakan dalam penelitian yang berfokus pada eksplorasi konsep dan fenomena yang kompleks.

Data dalam penelitian ini dikumpulkan dari berbagai sumber literatur primer dan sekunder. Literatur primer mencakup jurnal ilmiah yang memuat hasil penelitian terbaru terkait evaluasi pembelajaran dalam pendidikan Islam. Sedangkan literatur sekunder mencakup buku referensi dan dokumen resmi yang membahas prinsip-prinsip evaluasi, konsep reliabilitas alat ukur, serta faktor-faktor yang memengaruhi efektivitas evaluasi. Proses pengumpulan data dilakukan secara sistematis untuk memastikan bahwa informasi yang diperoleh relevan dan valid sesuai dengan tujuan penelitian. Menurut Zed studi pustaka memberikan keunggulan dalam menyediakan landasan teoretis yang kuat dan mendalam bagi penelitian.

Analisis data dilakukan secara deskriptif dengan tujuan untuk mengidentifikasi dan memahami berbagai aspek terkait evaluasi pembelajaran dalam pendidikan Islam. Peneliti menggambarkan konsep reliabilitas alat ukur, jenis-jenis evaluasi, serta faktor-faktor yang memengaruhi hasil evaluasi, seperti kemampuan evaluator, kondisi peserta didik, dan lingkungan pembelajaran. Melalui analisis deskriptif, penelitian ini mampu mengungkap hubungan antara berbagai faktor tersebut dan bagaimana mereka memengaruhi keberhasilan evaluasi dalam pendidikan Islam. Hal ini sejalan dengan pandangan Sugiyono yang menyatakan bahwa analisis deskriptif dapat memberikan gambaran yang jelas tentang fenomena yang diteliti tanpa mengubah konteks aslinya.

HASIL DAN PEMBAHASAN

A. Pengertian Reliabilitas

Reliabilitas adalah serangkaian pengukuran atau serangkaian alat ukur yang memiliki konsistensi bila pengukuran yang dilakukan dengan alat ukur itu dilakukan secara berulang. Reliabilitas diterjemahkan dari kata reliability yang berarti hal yang dapat dipercaya (tahan uji). Sebuah tes dikatakan mempunyai reliabilitas yang tinggi jika tes tersebut memberikan data hasil yang ajeg (tetap) walaupun diberikan pada waktu yang berbeda kepada responden yang sama. Reliabilitas merupakan salah satu kriteria yang harus dipenuhi oleh instrumen pengukuran sebelum digunakan untuk mengumpulkan data. Reliabilitas mengacu pada ketetapan atau kekonsistenan alat ukur, meskipun digunakan berulang kali pada subjek yang sama.

reliabilitas sebagai konsistensi sebuah hasil penelitian dengan menggunakan

berbagai metode penelitian dalam kondisi (tempat dan waktu) yang berbeda. Secara khusus, konsep reliabilitas mengacu pada konsistensi hasil score pada item-item yang terdapat pada kuesioner Anda sehingga uji reliabilitas sesungguhnya menguji ketepatan skala-skala pengukuran instrumen penelitian. Dengan demikian tujuan utama uji reliabilitas instrumen penelitian ialah untuk mengukur konsistensi alat ukur yang digunakan peneliti kuantitatif. Dalam konteks ini, peneliti hendak mengetahui apakah terdapat ketepatan hasil pengukuran pada sampel yang sama dalam waktu yang berbeda. Dengan kata lain, sebuah instrumen penelitian, misalnya kuesioner dinyatakan reliabel jika instrumen tersebut dapat menyediakan hasil skor yang konsisten pada setiap pengukuran. Dengan demikian, alat pengukuran tersebut (butir-butir pernyataan/pertanyaan) tetap menyediakan hasil pengukuran yang konsisten dalam waktu yang berbeda

Reliabilitas adalah konsistensi pengukuran dan stabilitas skor artinya pengukuran yang sama menghasilkan hasil yang sama. Jika dosen melakukan tes yang sama pada peserta yang sama pada kesempatan yang berbeda, dan hasilnya serupa, maka tes itu reliabel. Prinsip reliabilitas adalah konsistensi skor, instruksi yang jelas, dan pertanyaan yang jelas.

B. Jenis – jenis Reliabilitas

Reliabel adalah syarat yang harus terpenuhi dari suatu instrumen penelitian data penelitian. Instrumen penelitian harus memenuhi angka reliabel yang memadai. Menurut Djali and Muljono, (2008) reliabilitas terbagi menjadi:

1. Reliabilitas Konsistensi Tanggapan

Reliabilitas konsistensi tanggapan merupakan reliabilitas yang mengutamakan tanggapan responden atau objek penelitian, sudah masuk kategori baik atau konsisten. Hal ini berarti, hasil ukur dari instrumen penelitian tetap mendapatkan hasil yang sama jika dilakukan pengukuran berulang kali. Jika pada saat pengukuran dilakukan terjadi ketidaksesuaian hasil ukur dengan hasil ukur sebelumnya (tidak Konsisten) maka hasil pengukuran dianggap tidak menerangkan keadaan yang sebenarnya. Untuk mengetahui apakah instrumen penelitian tersebut telah baik atau konsisten, pengukuran selalu dilakukan terhadap objek yang sama. Dalam memeriksa reliabilitas suatu instrumen dapat dilakukan dengan tiga cara:

- a. Menggunakan Teknik test-retest yaitu dilakukan pengujian lebih dari satu kali terhadap instrument yang sama.
- b. Menggunakan Teknik bagi dua yaitu melakukan pengujian instrumen kepada dua kelompok item yang ekuivalen secara bersamaan.
- c. Menggunakan Teknik Bentuk ekuivalen yaitu dengan menggunakan dua instrumen yang ekuivalen pada waktu yang sama kepada responden.

2. Reliabilitas Konsistensi Gabungan

Item Reliabilitas konsistensi gabungan item merupakan reliabilitas yang menekankan kekonsistenan antara antara item-item suatu tes atau instrument. Jika terdapat ketidakkonsistenan pada hasil ukur antara item satu dengan yang lain dalam instrumen yang sama terhadap obyek yang sama, maka ada kesalahan pada instrumen atau alat ukur yang digunakan. Dengan demikian dapat dikatakan bahwa instrument tersebut tidak reliabel atau memiliki reliabilitas yang rendah. Menentukan reliabilitas konsistensi Gabungan item ini dapat dilakukan dengan pengelompokan instrumen berdasarkan jenis instrumennya pertama dengan Tes Objektif kedua Tes Uraian, dan ketiga Tes Afektif.

Secara garis besar ada dua macam cara menentukan reliabilitas instrumen, yaitu reliabilitas eksternal dan reliabilitas internal. Berikut penjabarannya:

1. Uji Reliabilitas eksternal

Reliabilitas eksternal merupakan reliabilitas yang apabila ukuran atau kriterianya

berada di luar instrumen. Di dalam reliabilitas eksternal, terdapat beberapa tes reliabilitas yang dapat digunakan, seperti:

- a. Metode tes ulang (test-retest method). Metode tes ulang dilakukan dengan mengujicobakan sebuah tes kepada sekelompok peserta didik sebanyak dua kali pada waktu yang berbeda. Skor hasil uji coba pertama dikorelasikan dengan skor hasil uji coba kedua dengan menggunakan teknik korelasi product momen
 - b. Metode Paralel. Metode paralel dilakukan dengan menguji cobakan dua buah instrumen yang dibuat hampir sama. Uji coba dilakukan terhadap sekelompok responden. Setiap responden mengerjakan atau mengisi kedua buah tes. Kemudian skorskor kedua buah tes tersebut dikorelasikan dengan teknik korelasi Product Moment. Angka korelasi ini menunjukkan tingkat reliabilitas instrumen. Metode paralel ini digunakan untuk mengatasi kelemahan yang terjadi pada metode tes ulang. Ketika dua tes yang digunakan ternyata berbeda, maka faktor carryover effect tidak menjadi masalah lagi, walaupun bisa saja faktor mengingat pada jawaban tes pertama sedikit berpengaruh pada tes kedua, khususnya apabila ditemukan soal yang benar-benar mirip atau bahkan sama.
 - c. Metode Belah dua. Metode belah dua digunakan untuk mengatasi kelemahankelemahan yang terjadi pada metode bentuk paralel dan metode tes ulang karena metode ini memungkinkan mengestimasi reliabilitas tanpa harus menyelenggarakan tes dua kali. Beberapa formula yang termasuk ke dalam metode belah dua, sebagai berikut Formula Spearman-Brown, Formula Flanagan, dan formula rulon.
2. Uji Reliabilitas Internal Pada reliabilitas internal, uji coba dilakukan hanya satu kali dan menggunakan satu instrumen. Kemudian hasil uji coba dianalisis dengan menggunakan rumus reliabilitas instrumen. Reliabilitas internal dapat dilihat diuji berdasarkan bentuk instrumennya, seperti:
- a. Reliabilitas instrumen bentuk dikotomi
 - 1) Rumus Kuder Richardson 20 (KR-20)

Reliabilitas untuk instrumen yang berbentuk dikotomi yaitu instrumen dengan pemberian skor 0 dan 1 maka pengujiannya dapat dilakukan dengan menggunakan rumus Kuder Richardson 20 (KR-20) dan Kuder Richardson 21 (KR-21). Penggunaan rumus KR. 20 digunakan apabila alternatif jawaban pada instrumen bersifat dikotomi, misalnya benar-salah dan pemberian skor = 1 dan 0.
 - 2) Rumus Kuder Richardson 21 (KR-21) Penggunaan rumus KR. 21 digunakan apabila alternatif jawaban pada instrumen bersifat dikotomi, misalnya benar-salah dan pemberian skor = 1 dan 0.
 - b. Reliabilitas instrumen bentuk kontinum Reliabilitas untuk instrumen yang berbentuk kontinum yaitu instrumen dengan pemberian skor yang skornya merupakan rentangan 0 – 10, 0 – 100 atau berbentuk skala 1 – 3, 1 – 5 atau 1-10, maka pengujiannya dapat dilakukan dengan menggunakan rumus Alpha Cronbach.

C. Cara Pengukuran Reliabilitas

Pengujian reliabilitas suatu instrumen bisa dilakukan secara eksternal maupun internal. Pengujian secara eksternal bisa dilakukan dengan test-retest (stability), equivalent, dan gabungan keduanya. Sedangkan pengujian secara internal reliabilitas instrumen dapat diuji dengan menganalisis konsistensi butir-butir yang ada pada instrumen dengan teknik tertentu. Berikut pengujian reliabilitas:

1. Test re-test Tes re-test berarti melakukan percobaan instrumen lebih dari satu kali kepada responden yang sama pada waktu yang berbeda. Koefisien korelasi yang dihasilkan dari setiap percobaan dijadikan sebagai acuan reliabilitas data. Jika

koefisien korelasi bernilai positif dan signifikan pada setiap percobaan maka dapat dikatakan reliabel.

2. Ekuivalen Ekuivalen maksudnya adalah terdapat pertanyaan/pernyataan yang memiliki makna yang sama meskipun menggunakan Bahasa yang berbeda. Cara ini perlu menggunakan dua instrumen yang ekuivalen kepada responden yang sama. Hasil korelasi antara kedua instrument dijadikan sebagai acuan reliabilitas data. Bila korelasi positif dan signifikan, maka instrumen dapat dinyatakan reliabel.
3. Gabungan Uji reliabilitas gabungan merupakan gabungan dari test-retest (stability) dan ekuivalen. I dilakukan dengan mengujikan dua buah instrumen yang ekuivalen terhadap responden yang sama dengan lebih dari satu kali pengujian. Selanjutnya, nilai korelasi antara kedua instrument pada pengujian pertama dan kedua yang kemudian dilakukan korelasi silang. Jika hasil semua pengujian menghasilkan koefisien korelasi yang positif dan signifikan, maka dapat dinyatakan bahwa instrumen itu reliabel.
4. Internal Consistency Internal consistency merupakan uji reliabilitas dengan hanya mengujicobakan instrumen satu kali. Data yang diperoleh kemudian dianalisis menggunakan teknik-teknik tertentu. Hasil analisis dijadikan sebagai acuan reliabilitas data. Teknik-teknik yang digunakan dapat berupa rumus-rumus statistik seperti Pengujian reliabilitas instrumen dapat dilakukan dengan teknik belah dua dari Spearman Brown , KR20, KR21 dan Anova Hoyt.

Salah satu hal mendasar yang perlu diperhatikan oleh setiap peneliti kualitatif ialah bagaimana mengukur akurasi dan konsistensi penelitian kualitatif. Berikut ini dijelaskan tentang beberapa strategi yang sudah digunakan secara luas dalam penelitian kualitatif khususnya dalam upaya meningkatkan reliabilitas penelitian.

1. Audit Proses Penelitian (Inquiry Audit)

Audit Proses Penelitian (Inquiry Audit) Lincoln dan Guba (1985) menekankan 'inquiry audit' (audit terhadap proses penelitian) untuk menilai sebuah reliabilitas penelitian kualitatif. Menurut para ahli Campbell, Clont dan Seal konsistensi data dapat diperoleh dari hasil penilaian terhadap data-data primer yang autentik, proses reduksi data melalui koding tematik dan analitik serta bukti hasil koding yang tersimpan dalam nodes. Syarat-syarat pengukuran tersebut sangat efisien dan efektif diperoleh melalui teknik manajemen dan analisis data dengan software NVivo karena semua proses penelitian terekam dalam NVivo dan dapat ditampilkan pada saat analisis data.

2. Audit Trail

Audit trail merupakan teknik yang dapat digunakan oleh peneliti untuk menunjukkan proses pengumpulan dan analisis data penelitian yang lengkap dan mendetail (Shenton, 2004). Dia menyarankan agar audit trail proses dan analisis data penelitian dapat dibuat dalam bentuk grafik k. Dalam grafik k-grafik k tersebut, peneliti dapat menunjukkan bagaimana data penelitian dapat menghasilkan rekomendasi berdasarkan tahapan-tahapan proses pengumpulan dan analisis data (data oriented audit trail).

3. Review Hasil Hasil Analisis Data

Dalam konteks penelitian yang dilakukan bersama-sama dengan tim peneliti, kita dapat meminta kolega peneliti untuk melakukan koding terhadap data yang sama. Namun dalam konteks penelitian tunggal yang hanya dilakukan oleh seorang peneliti, pencapaian konsistensi hasil analisis data dapat diperoleh dengan expert-review, yakni meminta ahli atau dosen pembimbing untuk memberikan penilaian terhadap hasil koding sesuai dengan tujuan penelitian.

Tujuan utama dari teknik pengujian reliabilitas koding kualitatif (Inter-Coder Reliability/ ICR) ialah untuk mengurangi bias subjektif dari peneliti. Beberapa ahli

menjelaskan bahwa inter-coder reliability (ICR) merupakan teknik pengujian reliabilitas yang paling umum dilakukan dalam penelitian kualitatif. ICR dalam penelitian kualitatif digunakan untuk menentukan konsistensi hasil koding sehingga menuntut adanya lebih dari satu koder atau rater. Konsistensi koding tersebut ditentukan berdasarkan tingkat persetujuan (agreement) dua atau lebih koder. Secara khusus perlu dipahami bahwa persetujuan dalam konteks ini berkaitan dengan kesamaan kedua koder (peneliti) untuk menentukan makna ungkapan informan kunci penelitian. Pemaknaan terhadap informasi yang disampaikan oleh para informan penelitian kualitatif merupakan suatu proses analitis yang dilakukan peneliti. Proses ini menghasilkan tema-tema dan subtema hasil analisis data yang disebut dengan codes.

Uji reliabilitas instrumen penelitian ialah untuk mengukur konsistensi alat ukur yang digunakan peneliti kuantitatif. Dalam penelitian kuantitatif, terdapat dua cara umum yang digunakan banyak peneliti untuk menentukan tingkatan reliabilitas:

1. Test-retest reliability

Pendekatan test-retest atau disebut juga pengukuran ulang, metode tes ulang, reliabilitas ulang-uji, tes ulang tes digunakan ketika seorang peneliti melakukan tes pada sampel penelitian yang sama dalam waktu yang berbeda. Dengan kata lain, seorang responden diberi kuesioner dan/atau tes yang sama dalam waktu yang berbeda, lalu hasil dari kedua tes tersebut dikorelasikan untuk melihat konsistensi hasil. Semakin tinggi tingkat korelasi pada tes pertama dan kedua, makin baik reliabilitas skala pengukuran Anda. Beberapa ahli tidak menyarankan Anda menggunakan tes ini karena pertimbangan waktu dan biaya. Menurut Pallant (2005) seorang peneliti yang menggunakan tes ini dapat saja mengalami risiko rendahnya korelasi antara tes pertama dan kedua karena faktor responden yang bersifat tidak stabil (labil). Kendatipun Sugiyono meyakini test-retest reliability sebagai sesuatu (stability test), sesungguhnya yang terjadi bukanlah sebuah stabilitas yang permanen tetapi hanya bersifat sementara karena situasi tes pertama dan kedua yang berbeda. Dalam konteks ini, menamakan uji reliabilitas test-retest ini dengan sebutan temporal stability.

2. Tes konsistensi internal (Internal consistency)

Sementara itu, tes konsistensi internal yang paling tepat dipakai ialah Alpha's Cronbach atau disebut juga dengan alpha coefficient. Rentangan nilai koefisien alpha berkisar antara 0 (tanpa reliabilitas) sampai dengan 1 (reliabilitas sempurna). Namun demikian, beberapa pakar menjelaskan bahwa semakin banyak item/ pernyataan butir skala pengukuran Anda makin tinggi tingkatan nilai reliabilitasnya. Dalam SPSS Statistik, pengujian konsistensi internal melalui Cronbach's analysis merupakan yang paling umum digunakan peneliti

Pada pengukuran konsistensi eksternal diperlukan dua kali pengadministrasian instrumen atau soal, sementara untuk mengukur konsistensi internal hanya diperlukan satu kali pengadministrasian instrumen. Jadi pada pengukuran konsistensi internal ini hanya diperlukan satu set soal. Beberapa cara pengukuran konsistensi internal ini adalah sebagai berikut:

- a. Prosedur Bagi-Dua (Split-Half Procedure) Prosedur bagi dua ini meliputi penilaian terhadap satu set soal yang dibagi dua (biasanya soal dengan nomor ganjil dan nomor genap atau awal dan akhir) dikerjakan secara terpisah oleh setiap orang. Selanjutnya dihitung koefisien korelasi untuk kedua belahan soal tersebut. Koefisien korelasi yang diperoleh menunjukkan derajat korelasi terhadap kedua belahan soal, dan oleh karena itu menggambarkan konsistensi internal dari tes tersebut.
- b. Pendekatan-Pendekatan Kuder-Richardson (Kuder-Richardson Approaches) Metode yang paling sering digunakan untuk menentukan konsistensi internal adalah

Pendekatan Kuder-Richardson, khususnya rumus KR20 dan KR21. Formula ini hanya memerlukan tiga buah informasi yaitu: jumlah item tes, rata-rata (Mean), dan standar deviasi (SD). Akan tetapi sebagai catatan bahwa KR21 dapat digunakan hanya jika diasumsikan bahwa item-item memiliki tingkat kesulitan yang sama (are of equal difficulty).

D. Faktor – Faktor Yang Mempengaruhi Reliabilitas

Tinggi rendahnya koefisien reliabilitas dipengaruhi oleh beberapa faktor. Beberapa pakar menyebutkan faktor-faktor yang mempengaruhi reliabilitas tes secara berbeda. Menurut Robert M Thorndike dalam bukunya berjudul *Measurement and Evaluation in Psychology and Education Seventh Edition*, faktor yang mempengaruhi reliabilitas adalah: 1) Variabilitas kelompok yang diberikan tes (Variability of the group on the trait the test measures); 2) Level kelompok yang diberikan tes (Level of the group on the trait the test measures); Panjang tes (Length of the test); Teknik atau rumus yang digunakan untuk menghitung reliabilitas (Operations used for estimating the reliability).

Sedangkan menurut Gilbert Sax, beberapa kondisi yang mempengaruhi koefisien reliabilitas, yaitu Objektivitas penskoran: Reliabilitas pemberi skor (Objectivity in scoring: Scorer Reliability); Variabilitas kelompok yang dites (The variability of the group tested); Jumlah item pada tes (The number of items on the test); dan Tingkat kesulitan tes (The difficulty level of the test).

Menurut Ebel koefisien reliabilitas akan bertambah besar untuk skor:

1. Dari tes yang lebih panjang daripada tes yang lebih pendek (from a longer test than from a shorter test)
2. Dari tes yang tersusun dari item yang lebih homogen daripada tes yang lebih heterogen (from a test composed of more homogeneous items than from a more heterogeneous test)
3. Dari tes yang memiliki daya pembeda lebih tinggi daripada yang mempunyai daya pembeda lebih rendah (from a test composed of more discriminating item than from a test composed of less discriminating items)
4. Dari tes yang tingkat kesulitannya sedang daripada tes yang terlalu sulit atau terlalu mudah (from a test whose items are of middle difficulty than from a test composed mainly of quite difficult or quite easy items)
5. Dari kelompok yang mempunyai jarak kemampuan yang lebar daripada yang berkemampuan homogen (from a group having a wide range of ability than from a group more homogeneous in ability)
6. Tes yang membutuhkan waktu terbatas daripada waktu yang lebih longgar (from a speeded test than from one all examinees can complete in the time available)

Menurut Crocker dan Algina menyebutkan bahwa faktor yang mempengaruhi reliabilitas tes antara lain panjang suatu tes, kecepatan, homogenitas belahan, dan tingkat kesukaran soal. Menurut Alen dan Yen dan Crocker dan Algina tingkat kesukaran itu memegang peranan yang paling dominan. Hasil penelitian Aiken menunjukkan bahwa pengaruh tingkat kesukaran memegang peranan yang paling besar pada koefisien reliabilitas. Hal ini disebabkan karena menyangkut variasi jumlah soal yang dapat dijawab benar. Semakin sukar soal-soal dalam perangkat tes akan semakin besar pula variasi skor yang diperoleh belahan. Dengan demikian maka akan semakin besar pula reliabilitas tes tersebut. Sebaliknya, semakin rendah tingkat kesukaran suatu soal semakin kecil pula reliabilitasnya. Untuk itu harus dihindari banyaknya terkaan yang dilakukan peserta tes dan diusahakan menyesuaikan pengetahuan peserta tes dengan materi tes yang akan diujikan kepada mereka. Penambahan panjang tes akan menaikkan koefisien reliabilitas sepanjang soal yang digunakan untuk menambah tes itu memiliki kualitas yang sama baik

dengan soal-soal lainnya.

Berdasarkan paparan di atas, dapat kita identifikasi hal apa saja yang mempengaruhi reliabilitas tes hasil belajar, yaitu:

1. Jumlah butir tes (panjang tes)

Umumnya semakin besar jumlah butir soal tes semakin tinggi pula reliabilitasnya. Hal ini terjadi karena semakin panjang tes (semakin banyak butir soal) sehingga semakin banyak perilaku yang terukur dengan lebih tepat. Kenaikan reliabilitas sebagai fungsi dari panjang tes dijelaskan oleh formula Spearman-Brown, yaitu:

r_n = indeks reliabilitas setelah ditambahkan soal

n = perkalian penambahan awal

r = indeks reliabilitas awal

Apabila tes memiliki banyak butir, maka reliabilitasnya dapat meningkat, asal penambahan butir tersebut mengacu pada pendiskripsian yang jelas terhadap variabel yang diukur. Jadi, dalam menyiapkan suatu tes perlu dipertimbangkan jumlah butirnya. Walaupun ditemukan semakin banyak butir berarti semakin tinggi reliabilitas, bukanlah berarti bahwa perlu disiapkan butir tes sebanyak-banyaknya, tetapi tetap berada pada batas dimensi dan indikator dari variabel yang diukur.

2. Variabilitas kelompok atau heterogenitas kelompok

Faktor lain yang berpengaruh terhadap reliabilitas instrument adalah heterogenitas kelompok. Heterogenitas kelompok diperlihatkan oleh besar kecilnya varians distribusi skor subjek pada variabel yang diungkap oleh tes yang bersangkutan. Bila tes yang bersangkutan mengungkap kemampuan bahasa arab maka heterogenitas kelompok diperlihatkan oleh besarnya variasi skor kemampuan berbahasa arab di antara siswa yang dites. Hal ini juga sesuai dengan rumus reliabilitas, yaitu:

Koefisien reliabilitas menunjukkan sejauhmana sebuah tes secara konsisten menempatkan siswa di tengah-tengah kelompoknya. Koefisien reliabilitas secara langsung dipengaruhi oleh penyebaran skor dalam kelompok yang diukur. Semakin besar penyebaran skor (artinya semakin kelompok bervariasi) maka semakin besar pula indeks reliabilitas yang diperoleh. Akan semakin sulit untuk mengukur perbedaan kemampuan siswa dalam kelompok yang homogen daripada dalam kelompok yang heterogen. Variabilitas kelompok ini berimplikasi pada penyebaran skor. Koefisien reliabilitas secara langsung dipengaruhi oleh penyebaran skor dalam kelompok yang diukur. Hal-hal lain menjadi sama, semakin besar penyebaran skor maka semakin besar pula indeks reliabilitas yang diperoleh.

3. Objektivitas penskoran

Reliabilitas pemberi skor merujuk kepada keberadaan observer yang berbeda atau kesepakatan pemberi nilai yang mereka berikan pada instrument yang sama. Semakin tinggi kesepakatan, maka reliabilitas pemberi skor akan semakin tinggi juga. Begitu juga sebaliknya. Sebagian besar tes bakat dan tes prestasi standar mempunyai objektivitas yang tinggi. Butir-butir skor tes objektif seperti pilihan ganda dan skor yang dihasilkan tidak dipengaruhi oleh keputusan dan pendapat pemberi skor. Semakin tinggi tingkat objektivitas tes semakin tinggi pula tingkat reliabilitasnya.

4. Metode Estimasi Reliabilitas

Saat menguji koefisien reliabilitas tes standar, memutuskan metode yang digunakan untuk menentukan besarnya koefisien reliabilitas merupakan hal yang penting. Secara umum, besarnya koefisien reliabilitas berkaitan erat dengan metode yang digunakan untuk estimasi reliabilitas. Berikut ini beberapa tes untuk menentukan koefisien reliabilitas:

- a. Metode tes ulang (Test Retest Method): mungkin hasilnya lebih besar dibandingkan dengan metode belah dua jika interval waktunya pendek. Koefisien reliabilitas yang

- dihasilkan menjadi lebih kecil jika interval waktu tesnya ditingkatkan.
- b. Tes sejajar (Equivalent Test) tanpa waktu interval: Koefisien reliabilitas cenderung lebih rendah dibandingkan dengan metode belah dua atau tes ulang yang menggunakan interval waktu singkat.
 - c. Tes Sejajar dengan interval waktu: koefisien menjadi lebih kecil seiring dengan peningkatan interval waktu tes.
 - d. Metode belah dua (Split-half Method): Metode ini menyediakan sebuah indikasi konsistensi internal tes. Guru dapat menggunakan salah satu teknik analisis untuk estimasi reliabilitas, yang perlu ditekankan adalah penggunaan teknik tertentu memerlukan data pada kategori interval atau nominal
5. Level Kelompok Dan Tingkat Kesulitan Tes

Selain variabilitas dalam kelompok, level kelompok juga mungkin dapat berpengaruh dalam koefisien reliabilitas. Presisi pengukuran dari sebuah tes bisa berhubungan dengan level kemampuan dari orang yang diukur; bagaimanapun juga, tidak ada aturan yang dapat memformulasikan keadaan hubungan ini. Sifat hubungan ini bergantung pada bagaimana cara tes itu disusun. Untuk mereka yang merasa bahwa tes terlalu sulit, sepertinya mereka akan sering menebak, akurasinya akan menjadi rendah. Pada keadaan lain, jika tes terlalu mudah untuk kelompok itu, seperti mereka pasti bisa menjawab kebanyakan item dengan benar, tes tersebut bisa menjadi tidak efektif dalam membedakan dalam anggota kelompok. Hal ini ekuivalen dengan tingkat kesulitan tes. Tes yang sangat mudah atau sangat sulit tidak dapat mengukur perbedaan individu.

6. Homogenitas Tes

Homogenitas tes juga berpengaruh pada reliabilitas. Seratus item tes sejarah kelas X akan menjadi lebih reliabel daripada 100 item tes yang mengukur tes sejarah pada tingkat SMA. Begitu juga pada mata pelajaran lainnya, seperti matematika atau bahasa arab, yang lebih membutuhkan pengorganisasian yang lebih padat, dengan ketergantungan yang erat pada bukti-bukti, aturan-aturan, kemampuan, dan hasil, daripada seperti pada mata pelajaran sejarah. Ini adalah aspek lain dari homogenitas tes yang membuat reliabilitas tes yang tinggi menjadi lebih mudah dicapai pada tes matematika dan bahasa asing dibandingkan pada tes hasil belajar materi pendidikan lainnya.

KESIMPULAN

Reliabilitas alat ukur evaluasi dalam pendidikan Islam memegang peranan penting dalam memastikan validitas hasil evaluasi yang diperoleh. Reliabilitas mencerminkan konsistensi alat ukur dalam memberikan hasil yang stabil dan akurat ketika digunakan dalam berbagai situasi. Dalam konteks pendidikan Islam, alat ukur yang reliabel tidak hanya membantu mengevaluasi pencapaian pembelajaran secara objektif tetapi juga memastikan integrasi nilai-nilai Islam dalam proses evaluasi. Hasil evaluasi yang valid dan reliabel memungkinkan pendidik untuk memahami kemajuan peserta didik secara menyeluruh, baik dari segi akademik maupun spiritual, sehingga dapat dijadikan dasar untuk merancang langkah pembelajaran berikutnya.

Untuk mencapai reliabilitas yang tinggi, evaluator perlu memiliki kompetensi yang memadai dalam merancang dan menggunakan alat ukur yang relevan dengan tujuan pembelajaran. Kompetensi ini mencakup pemahaman tentang prinsip-prinsip evaluasi, pemilihan metode yang sesuai, serta kemampuan menganalisis dan menginterpretasikan hasil evaluasi. Institusi pendidikan Islam perlu berperan aktif dalam mendukung peningkatan kompetensi pendidik melalui pelatihan berkala tentang teknik evaluasi yang efektif dan terkini. Selain itu, lingkungan pembelajaran yang kondusif juga harus diperhatikan, karena suasana belajar yang mendukung akan membantu peserta didik

memberikan hasil terbaik selama proses evaluasi. Dengan kombinasi antara kompetensi evaluator yang baik dan lingkungan pembelajaran yang optimal, evaluasi dalam pendidikan Islam dapat menjadi instrumen yang efektif untuk mendukung pencapaian tujuan pendidikan secara menyeluruh.

DAFTAR PUSTAKA

- Abdul Muin, Buku Ajar Metode Penelitian Kuantitatif, (Malang: CV Literasi nusantara Abadi, 2023).
- Abigail Susana dkk, Metodologi Penelitian Kuantitatif, (Tanpa Kota: Yayasan Kita Menulis, 2023).
- Adhi Setiawan, Jurnal An Nur, Faktor – faktor yang Mempengaruhi Reliabilitas Tes, Vol.6, No.2, 2014.
- Creswell, J. W. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th Edition). California: SAGE Publications.
- Dyah Budiastuti & Agustinus Bandur, *Validitas Dan Reliabilitas Penelitian Dengan Analisis dengan NVIVO, SPSS dan AMOS*, (Jakarta: Mitra Wacana Media, 2018).
- Muhammad Fakhi Ramadhan, Rusydi A. Siroj, dan Muhammad win Afgani, *Journal on Education, Validitas dan Reliabilitas*, Vol.6, No.2, 2024.
- Nurgiyantoro, B. (2012). *Penilaian dalam Pengajaran Bahasa dan Sastra*. Yogyakarta: BPFE.
- Slamet Widodo, *Buku Ajar Metode Penelitian*, (Pangkal Pinang: CV Science Techno Direct, 2023).
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Suparman. (2020). *Evaluasi Pembelajaran dalam Perspektif Pendidikan Islam*. Jakarta: RajaGrafindo Persada.
- Tri Jampi Setiyorini, Zyah Rochmad Jaelani, dan Abdul Ngafif, *Didaktif: Jurnal Pendidikan dan Ilmu Pengetahuan, Analisis Faktor – faktor yang Mempengaruhi Reliabilitas Tes Tata Bahasa Inggris di Universitas Indonesia*, Vol.22, No.3, 2022.
- Zed, M. (2004). *Metode Penelitian Kepustakaan*. Jakarta: Yayasan Obor Indonesia.