



STIT Darul Hijrah Martapura, Kalimantan Selatan, Indonesia
TARBAWI: Jurnal Pendidikan dan Keagamaan

p-ISSN: 2460-1101, e-ISSN: 2775-3395
Vol. 14 No. 01 Juni 2026

KAJIAN KONSEPTUAL METODE BELAH DUA (*SPLIT-HALF*) DAN BENTUK PARALEL (*EQUIVALENCE*) DALAM EVALUASI PEMBELAJARAN PAI

Riduan Iksan¹, Dina Hermina², Mar-a Rabbi Radliyya³

^{1,2} Universitas Islam Negeri Antasari Banjarmasin, Indonesia

³ STIT Darul Hijrah Martapura, Indonesia

¹riduaniksan.uin2@gmail.com, ²dinahermina@uin-antasari.ac.id, ³mararadliyya@gmail.com

Abstract

Test reliability is a fundamental requirement in organizing educational evaluation instruments, especially in Islamic Religious Education (PAI), to ensure consistency and fairness in student assessment. This study aims to conceptually examine test reliability through the split-half method and parallel forms (equivalence) as well as its relevance to Islamic educational values. This study uses a qualitative approach with a literature review method. Data collection was carried out by documenting and analyzing secondary sources in the form of books, scientific articles, and related references. The literature selection was limited to publications from 2018–2025 sourced from national and international indexed journals and standard textbooks on psychometrics and educational evaluation. The results of the study show that the split-half method is effective for measuring internal consistency in a single test, where its correlation is corrected using the Spearman-Brown formula. Meanwhile, the parallel form method emphasizes cross-form consistency by arranging two equivalent test packages, which require a high level of test developer competence to minimize measurement error and prevent practice effects. In the context of PAI learning, the application of both reliability methods is not only an academic requirement but also reflects Islamic moral values such as justice, thoroughness, accountability, and professionalism as emphasized in the Qur'an and Hadith. PAI teachers are expected to improve their technical competence in statistical evaluation to produce objective and accountable learning assessments.

Keywords: Reliability; Split-Half Method; Parallel Forms; Educational Evaluation; PAI Learning

Abstrak

Reliabilitas tes merupakan syarat fundamental dalam penyelenggaraan instrumen evaluasi pendidikan, khususnya dalam Pendidikan Agama Islam (PAI), guna menjamin konsistensi dan keadilan penilaian peserta didik. Penelitian ini bertujuan untuk mengkaji secara konseptual reliabilitas tes melalui metode belah dua (split-half) dan bentuk paralel (equivalence) serta relevansinya dengan nilai-nilai pendidikan Islam. Penelitian ini menggunakan pendekatan kualitatif dengan metode

studi kepustakaan (literature review). Pengumpulan data dilakukan dengan mendokumentasikan serta menganalisis sumber-sumber sekunder berupa buku, artikel ilmiah, dan referensi terkait. Pemilihan literatur dibatasi pada publikasi rentang tahun 2018–2025 yang bersumber dari jurnal nasional dan internasional terindeks serta buku teks standar psikometri dan evaluasi pendidikan. Hasil penelitian menunjukkan bahwa metode belah dua efektif untuk mengukur konsistensi internal dalam satu kali pengujian tes, di mana korelasinya dikoreksi menggunakan rumus Spearman-Brown. Sementara itu, metode bentuk paralel menekankan konsistensi lintas-bentuk dengan menyusun dua paket tes yang setara, yang menuntut kompetensi tinggi dari penyusun tes untuk meminimalkan galat pengukuran serta mencegah efek belajar. Dalam konteks pembelajaran PAI, penerapan kedua metode reliabilitas tersebut bukan sekadar tuntutan akademik, melainkan juga merefleksikan nilai-nilai moral Islam seperti keadilan, ketelitian, tanggung jawab, dan profesionalisme sebagaimana ditekankan dalam Al-Qur'an dan Hadis. Guru PAI diharapkan meningkatkan kompetensi teknisnya dalam statistik evaluasi demi menghasilkan penilaian pembelajaran yang objektif dan dapat dipertanggungjawabkan.

Kata Kunci: Reliabilitas; Metode Belah Dua; Bentuk Paralel; Evaluasi Pendidikan; Pembelajaran PAI

PENDAHULUAN

Evaluasi merupakan salah satu komponen vital dan integral dalam keseluruhan sistem pendidikan, termasuk dalam pelaksanaan pembelajaran Pendidikan Agama Islam (PAI). Melalui proses evaluasi yang terstruktur, seorang pendidik dapat mengidentifikasi sejauh mana tujuan pembelajaran yang telah dirumuskan dapat tercapai secara optimal, baik pada ranah kognitif, afektif, maupun psikomotorik. Keberhasilan suatu proses pembelajaran tidak dapat dilepaskan dari kualitas instrumen evaluasi yang digunakan. Instrumen evaluasi yang bermutu tinggi wajib memenuhi dua syarat utama yang diakui dalam metodologi pengukuran, yaitu validitas dan reliabilitas.¹ Reliabilitas menempati posisi krusial karena berkaitan erat dengan tingkat konsistensi, keajegan, dan kestabilan hasil pengukuran apabila tes tersebut diujikan kembali pada kondisi atau waktu yang berbeda terhadap subjek yang sama.² Tanpa adanya reliabilitas yang teruji, data skor yang diperoleh dari hasil evaluasi tidak akan mampu mencerminkan kemampuan riil dari peserta didik secara akurat.

Dalam konteks pembelajaran Pendidikan Agama Islam, urgensi reliabilitas tes menjadi berlipat ganda karena materi yang disampaikan tidak sekadar berfokus pada transfer pengetahuan teoritis, melainkan juga menyangkut penanaman pemahaman nilai-nilai keagamaan, akidah, dan syariat yang mendalam. Penggunaan instrumen evaluasi yang tidak reliabel atau fluktuatif akan memicu bias penilaian, menimbulkan

¹Zainal Arifin, *Evaluasi Pembelajaran* (Bandung: Remaja Rosdakarya, 2021), 12–15.

²David L. Streiner, Geoffrey R. Norman, & John Cairney, *Health Measurement Scales: A Practical Guide to Their Development and Use*, 5th ed. (Oxford: Oxford University Press, 2019), 167–172.

ketidakadilan bagi peserta didik, dan mengakibatkan terjadinya kesalahan dalam pengambilan keputusan instruksional oleh guru.³ Oleh sebab itu, pemahaman mendalam mengenai metode pengujian reliabilitas instrumen menjadi suatu keharusan teoretis dan praktis bagi pendidik. Terdapat beberapa variasi metode yang diakui secara luas untuk mendeteksi koefisien reliabilitas, di antaranya adalah metode belah dua (*split-half method*) yang menitikberatkan pada konsistensi internal, serta metode bentuk paralel (*parallel forms* atau *equivalence*) yang berorientasi pada kesetaraan lintas-instrumen.

Meskipun studi mengenai reliabilitas instrumen telah banyak dipublikasikan, masih terdapat kesenjangan analisis (*gap analysis*) yang cukup signifikan. Kajian Kaplan & Saccuzzo (2018) tentang pengujian psikologis, misalnya, menjelaskan prosedur teknis reliabilitas secara rinci namun tidak mengintegrasikannya dengan konteks nilai atau orientasi pendidikan berbasis agama.⁴ Demikian pula Mertler (2020) dalam bukunya tentang penelitian pendidikan, yang mengakui bahwa sebagian besar literatur evaluasi pendidikan umum masih berpusat pada prosedur matematis tanpa menyentuh dimensi moral atau spiritual dari proses penilaian.⁵ Lebih lanjut, Streiner dkk. (2019) dalam ulasan komprehensif mereka tentang skala pengukuran kesehatan dan pendidikan secara eksplisit mencatat bahwa literatur psikometri Barat tidak mengakomodasi perspektif moral atau nilai keagamaan dalam pembahasan reliabilitas instrumen.⁶ Perbedaan mendasar artikel ini dengan kajian-kajian tersebut terletak pada pendekatan komprehensif yang memadukan secara mendalam antara teknis metodologis-statistik pengujian reliabilitas khususnya komparasi metode belah dua dan bentuk paralel dengan landasan moral-aksiologis pendidikan Islam yang bersumber dari Al-Qur'an dan Hadis.

Berdasarkan rumusan permasalahan tersebut, penelitian ini bertujuan untuk menjelaskan secara mendalam konsep dasar reliabilitas tes, menjabarkan prosedur matematis serta karakteristik operasional dari metode belah dua dan metode bentuk paralel, serta mengelaborasi relevansi teknis pengujian tersebut dengan nilai-nilai keadilan dan profesionalisme dalam perspektif pendidikan Islam. Kontribusi teoretis artikel ini adalah menjadi kajian pertama yang secara sistematis mengintegrasikan analisis komparatif antara dua metode reliabilitas (belah dua dan bentuk paralel) dengan kerangka aksiologis-teologis pendidikan Islam sebuah sintesis yang belum dijumpai pada artikel-artikel sebelumnya yang umumnya membahas salah satu aspek secara terpisah.

³Anas Sudijono, *Pengantar Evaluasi Pendidikan* (Jakarta: RajaGrafindo Persada, 2020), 208–210.

⁴Robert M. Kaplan & Dennis P. Saccuzzo, *Psychological Testing: Principles, Applications, and Issues*, 9th ed. (Boston: Cengage Learning, 2018), 98–105.

⁵Craig A. Mertler, *Introduction to Educational Research*, 2nd ed. (Los Angeles: SAGE, 2020), 88–92.

⁶David L. Streiner, Geoffrey R. Norman, & John Cairney, *Health Measurement Scales* (Oxford: Oxford University Press, 2019), 170.

Sistematika susunan artikel ini disajikan secara bertahap untuk mempermudah pemahaman pembaca. Setelah bagian pendahuluan ini, bagian kedua akan menguraikan metode penelitian yang digunakan untuk membedah masalah. Bagian ketiga menyajikan hasil dan pembahasan yang mencakup definisi reliabilitas, analisis mendalam beserta prosedur perhitungan metode belah dua, serta pemaparan komprehensif mengenai karakteristik dan teknis statistika metode bentuk paralel yang diintegrasikan dengan perspektif Al-Qur'an dan Hadis.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kualitatif dengan metode studi kepustakaan (library research) untuk mengkaji konsep reliabilitas tes, khususnya metode belah dua dan bentuk paralel, serta integrasinya dengan nilai-nilai pendidikan Islam.⁷ Studi kepustakaan dipilih karena fokus penelitian ini diarahkan pada analisis teoretis, konseptual, dan sintesis kritis terhadap literatur-literatur ilmiah yang relevan, sehingga peneliti berperan sebagai instrumen utama yang mengumpulkan, mereduksi, menganalisis, dan menginterpretasikan data secara objektif.⁸

Sumber data dalam penelitian ini adalah bahan kajian kepustakaan sekunder yang terpercaya, mencakup buku-buku teks standar evaluasi pendidikan dan psikometri, jurnal ilmiah nasional dan internasional yang terindeks, serta teks primer keagamaan berupa ayat-ayat Al-Qur'an dan matan Hadis. Prosedur seleksi literatur menerapkan kriteria inklusi sebagai berikut: (a) rentang tahun publikasi 2018–2025 untuk sumber jurnal dan buku teks, dengan pengecualian pada karya klasik yang bersifat foundational seperti Zed (2008) dan sumber keagamaan primer; (b) sumber yang membahas secara langsung topik reliabilitas instrumen, evaluasi pendidikan, atau integrasi nilai Islam dalam pengukuran pendidikan; (c) diterbitkan dalam bahasa Indonesia atau bahasa Inggris. Kriteria eksklusi meliputi sumber yang tidak tersedia teks lengkapnya, sumber opini tanpa landasan empiris atau teoritis yang jelas, serta artikel yang tidak relevan dengan psikometri atau evaluasi pendidikan keagamaan.

Metode pengumpulan data dilakukan melalui teknik dokumentasi, yaitu dengan mengidentifikasi, mencatat, dan mengklasifikasikan informasi teks yang berkaitan dengan rumus Spearman-Brown, korelasi Product Moment, teori tes klasik (Classical Test Theory), serta konsep keadilan (al-'adl) dan ketelitian (itqan) dalam penilaian.

Teknik analisis data yang diterapkan adalah analisis isi (content analysis) dan analisis komparatif. Prosedur analisis isi dilakukan melalui empat tahap sistematis: (1) unitisasi, yakni mengidentifikasi unit-unit teks yang relevan dengan tema reliabilitas dari setiap sumber; (2) kategorisasi, yakni mengelompokkan unit teks ke dalam

⁷Sugiyono, *Metode Penelitian Kualitatif* (Bandung: Alfabeta, 2018), 9.

⁸Mestika Zed, *Metode Penelitian Kepustakaan* (Jakarta: Yayasan Obor Indonesia, 2008), h. 3.

kategori tema (konsep dasar reliabilitas, prosedur metode belah dua, prosedur metode bentuk paralel, dan nilai Islam); (3) inferensi, yakni menarik kesimpulan konseptual dari pola yang teridentifikasi; dan (4) verifikasi, yakni memvalidasi inferensi melalui triangulasi sumber. Peneliti mengurai secara mendalam formula matematis dari masing-masing metode reliabilitas, menganalisis kelebihan dan kekurangannya, kemudian mengomparasikan karakteristik operasionalnya. Untuk memastikan validitas hasil penelitian, dilakukan triangulasi sumber data dengan membandingkan pandangan berbagai pakar evaluasi pendidikan, serta pengujian melalui diskusi sejawat guna menjaga konsistensi dan akurasi interpretasi teoretis.

HASIL DAN PEMBAHASAN

A. Konsep Dasar Reliabilitas

Reliabilitas secara etimologis diadopsi dari istilah Inggris, yaitu *reliability*, yang mengandung makna keandalan, kestabilan, atau konsistensi.⁹ Dalam ranah evaluasi pendidikan, reliabilitas didefinisikan sebagai indeks yang menunjukkan sejauh mana suatu alat pengukur atau instrumen tes dapat dipercaya dan diandalkan untuk menghasilkan skor yang konsisten apabila pengukuran dilakukan secara berulang-ulang terhadap subjek yang sama pada kondisi yang setara. Suatu tes dikategorikan memiliki reliabilitas yang tinggi apabila hasil pengukuran yang diperoleh menunjukkan tingkat keajegan yang konstan atau tidak mengalami perubahan yang signifikan meskipun diujikan dalam lintasan waktu yang berbeda.¹⁰

Secara metodologis, manfaat dari pengujian reliabilitas instrumen penilaian memegang peranan yang sangat fundamental dalam menjamin kualitas hasil evaluasi secara keseluruhan. Data evaluasi yang memiliki reliabilitas tinggi berfungsi sebagai basis data empiris yang kredibel bagi pendidik dalam memetakan profil instruksional secara holistik. Sebaliknya, instrumen yang tidak stabil atau tidak reliabel akan menghasilkan skor yang penuh dengan galat (error), yang pada akhirnya mencederai prinsip keadilan dalam penilaian pendidikan.¹¹

B. Jenis-jenis Reliabilitas

Dalam psikometri dan evaluasi pendidikan, dikenal beberapa jenis reliabilitas yang masing-masing memiliki fokus pengukuran dan prosedur pengujian yang berbeda, namun semuanya bertujuan untuk memastikan bahwa alat ukur memberikan hasil

⁹Derri Adi Fernando dan Fajri Ismail, "Pentingnya Validitas dan Reliabilitas Instrument Evaluasi pada Pembelajaran Pendidikan Agama Islam," *TARBAWI: Jurnal Pendidikan dan Keagamaan* 11, no. 01 (2023): 45–58.

¹⁰UIN Maulana Malik Ibrahim dan Vick Ainun Haq, "Menguji Validitas dan Reliabilitas pada Mata Pelajaran Al Qur'an Hadits Menggunakan Korelasi Produk Moment-Spearman Brown," *An-Nawa: Jurnal Studi Islam* 4, no. 1 (2022): 78–92. <https://doi.org/10.37758/annawa.v4i1.419>

¹¹Fityatul Muharromah, Farid Al Fani, dan Ahmad Rosyid Ridho, "Reliabilitas Alat Ukur: Jenis-Jenis, Cara Pengukuran, dan Faktor-Faktor yang Memengaruhinya dalam Pembelajaran," *Jurnal Edukasi* 13, no. 01 (2025): 12–25.

yang stabil dan dapat dipercaya.¹² Pemilihan jenis reliabilitas harus disesuaikan dengan karakteristik instrumen dan tujuan penilaian yang hendak dicapai.

Setidaknya terdapat empat jenis reliabilitas yang paling umum digunakan dalam konteks evaluasi pembelajaran:

1. **Reliabilitas Test-Retest (Stabilitas).** Jenis ini mengukur konsistensi hasil tes yang diberikan kepada kelompok yang sama pada dua waktu yang berbeda. Koefisien korelasinya menunjukkan stabilitas skor dari waktu ke waktu. Kelemahan utamanya adalah adanya efek ingatan (memory effect) pada testee jika jeda waktu terlalu singkat.
2. **Reliabilitas Antar-Rater (Inter-Rater Reliability).** Jenis ini mengukur konsistensi penilaian antara dua penilai atau lebih terhadap respons atau kinerja yang sama. Relevan untuk instrumen seperti rubrik esai, observasi, atau penilaian kinerja dalam evaluasi PAI.
3. **Reliabilitas Konsistensi Internal (Alpha Cronbach dan Kuder-Richardson).** Jenis ini mengukur sejauh mana seluruh butir soal dalam sebuah tes secara konsisten mengukur konstruk yang sama. Koefisien Alpha Cronbach paling sering digunakan untuk soal berskala politomi, sedangkan Kuder-Richardson (KR-20 dan KR-21) digunakan untuk soal dikotomi (benar-salah).
4. **Reliabilitas Belah Dua dan Bentuk Paralel.** Kedua jenis inilah yang menjadi fokus kajian artikel ini dan akan dibahas secara mendalam pada sub-bahasan berikutnya.

C. Metode Belah Dua (Split-Half Reliability)

Metode belah dua (*split-half reliability*) merupakan salah satu teknik pengujian yang digunakan untuk mengukur koefisien konsistensi internal dari suatu instrumen tes. Karakteristik utama dari metode ini adalah pelaksanaannya yang praktis, sebab peneliti atau guru hanya perlu mengujicobakan satu perangkat tes sebanyak satu kali (*single-trial administration*) kepada sekelompok responden. Prosedur operasional pengujian ini dimulai dengan membagi seluruh butir soal dalam tes tersebut menjadi dua bagian yang setara, menggunakan pendekatan ganjil-genap.¹³

Setelah pengelompokan skor selesai, langkah selanjutnya adalah menghitung koefisien korelasi antara skor bagian pertama dan bagian kedua menggunakan rumus korelasi Product Moment dari Karl Pearson:

$$r_{12} = (n \times \Sigma XY - \Sigma X \times \Sigma Y) / \sqrt{[n \times \Sigma X^2 - (\Sigma X)^2] \times [n \times \Sigma Y^2 - (\Sigma Y)^2]}$$

Keterangan: X merupakan skor total bagian pertama (ganjil), Y merupakan skor total bagian kedua (genap), dan n menyatakan jumlah total responden atau peserta tes. Koefisien korelasi (r_{12}) yang dihasilkan barulah mencerminkan reliabilitas dari setengah

¹²Saifuddin Azwar, Reliabilitas dan Validitas (Yogyakarta: Pustaka Pelajar, 2022), 45–50.

¹³David L. Streiner, Geoffrey R. Norman, & John Cairney, Health Measurement Scales: A Practical Guide to Their Development and Use, 5th ed. (Oxford: Oxford University Press, 2019), 167–172.

tes. Oleh karena itu, untuk mengestimasi koefisien reliabilitas keseluruhan perangkat tes, koefisien r_{12} wajib dikoreksi menggunakan formula Spearman-Brown:

$$r_{SB} = (2 \times r_{12}) / (1 + r_{12})$$

Sebagai simulasi perhitungan, apabila diperoleh nilai korelasi antar-belahan (r_{12}) sebesar 0,70, maka:

$$r_{SB} = (2 \times 0,70) / (1 + 0,70) = 1,40 / 1,70 \approx 0,82$$

Nilai koefisien 0,82 ini menunjukkan tingkat reliabilitas yang tinggi, karena berada di atas ambang batas standar pengujian instrumen pendidikan (umumnya $\geq 0,70$), sehingga perangkat soal tersebut dinilai sangat layak dan dapat dipercaya untuk digunakan dalam evaluasi pembelajaran.¹⁴

Dalam konteks pembelajaran PAI, implementasi metode belah dua sangat relevan untuk memastikan bahwa seluruh butir soal benar-benar homogen dan mengukur domain kompetensi yang sama (konsistensi internal). Ketelitian dalam menyusun dan menguji konsistensi internal ini merupakan manifestasi dari nilai profesionalisme dan ketelitian ilmiah yang sejalan dengan ajaran Islam. Kajian komparatif terhadap berbagai metode estimasi reliabilitas termasuk Cronbach Alpha, KR-20, KR-21, dan Split-Half menunjukkan bahwa masing-masing formula memiliki keunggulan tersendiri bergantung pada karakteristik instrumen yang diujikan, sehingga pemilihannya harus disesuaikan dengan konteks evaluasi pembelajaran.¹⁵ Allah Swt. berfirman dalam Surah Az-Zalzalah ayat 7–8:

فَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ خَيْرًا يَرَهُ 7 (وَمَنْ يَعْمَلْ مِثْقَالَ ذَرَّةٍ شَرًّا يَرَهُ) 8

"Siapa yang mengerjakan kebaikan seberat zarrah, dia akan melihat (balasan)-nya. Siapa yang mengerjakan kejahatan seberat zarrah, dia akan melihat (balasan)-nya." (QS. Az-Zalzalah: 7–8)

Ayat mulia ini memberikan landasan aksiologis yang kokoh bahwa dalam pandangan Islam, segala bentuk penilaian dilakukan dengan tingkat akurasi yang luar biasa cermat, adil, dan tidak ada satu pun aspek terkecil yang terabaikan. Konsep keadilan ilahiah ini menuntut para pendidik PAI untuk memastikan bahwa alat evaluasi hasil belajar yang mereka gunakan memiliki keajegan yang teruji secara ilmiah.

D. Metode Bentuk Paralel (Parallel Forms / Equivalence)

Metode pengujian reliabilitas selanjutnya adalah metode bentuk paralel, yang dikenal juga dengan istilah *parallel forms* atau uji ekuivalensi. Berbeda dengan metode belah dua yang menguji konsistensi internal dalam satu kali tes, metode bentuk paralel bertujuan untuk menguji konsistensi lintas-instrumen. Pendekatan ini dilakukan dengan cara menyusun dua perangkat tes yang berbeda secara redaksional namun

¹⁴Craig A. Mertler, *Introduction to Educational Research*, 2nd ed. (Los Angeles: SAGE, 2020), 88–92.

¹⁵Hari Sugiharto Setyaedhi, "Comparative Test of Cronbach's Alpha Reliability Coefficient, KR-20, KR-21, and Split-Half Method," *Journal of Education Research and Evaluation* 8, no. 1 (2024): 47–57. <https://doi.org/10.23887/jere.v8i1.68164>.

memiliki kesetaraan pada ranah substansi materi, tujuan pengukuran, jumlah butir soal, tingkat kesukaran, dan karakteristik psikometrisnya.¹⁶

Secara teoritis, pengujian metode bentuk paralel berakar pada Teori Tes Klasik (*Classical Test Theory*) yang mengasumsikan bahwa skor tampak (*observed score*) seseorang merupakan akumulasi dari skor sejati (*true score*) dan komponen galat pengukuran (*error score*). Dua buah instrumen tes dikategorikan paralel sejati apabila keduanya mengukur konstruk psikologis yang sama, memiliki nilai ekspektasi skor sejati yang identik, serta memiliki varians galat pengukuran yang setara.¹⁷

Prosedur teknis dalam implementasi metode bentuk paralel menuntut tahapan sistematis yang ketat, yang disusun mengikuti urutan berikut:

1. **Penyusunan Kisi-kisi Tunggal.** Penyusun tes wajib merumuskan satu cetak biru (kisi-kisi) soal yang sama, yang menetapkan secara detail indikator kompetensi, cakupan materi pembelajaran, dan distribusi tingkat kesulitan butir soal.
2. **Pengembangan Dua Paket Tes.** Penulis soal mengembangkan butir-butir pertanyaan secara terpisah untuk menghasilkan dua paket tes (Paket A dan Paket B) dengan mengacu secara ketat pada kisi-kisi tunggal tersebut.
3. **Pelaksanaan Uji Coba.** Kedua perangkat tes diujicobakan kepada kelompok sampel yang representatif.
4. **Analisis Statistik Komparatif.** Melakukan analisis statistik komparatif terhadap karakteristik butir soal untuk memastikan kesetaraan rerata (mean), simpangan baku (standard deviation), serta indeks daya beda dari kedua paket tes.
5. **Penghitungan Koefisien Korelasi.** Menghitung koefisien korelasi skor total antara kedua perangkat tes menggunakan rumus korelasi Product Moment:

$$r = \text{cov}(A,B) / (\sigma_A \times \sigma_B)$$

Keterangan: $\text{cov}(A,B)$ adalah nilai kovarians antara skor tes Paket A dan Paket B, yang dihitung dengan rumus:

$$\text{cov}(A,B) = [\sum(A_i - \bar{A})(B_i - \bar{B})] / (n - 1)$$

dengan A_i dan B_i adalah skor individu ke- i pada Paket A dan B, $\bar{A}$ dan $\bar{B}$ adalah rerata masing-masing paket, dan n adalah jumlah responden. σ_A dan σ_B masing-masing merepresentasikan standar deviasi dari skor tes Paket A dan Paket B. Koefisien korelasi yang dihasilkan langsung menjadi indikator koefisien reliabilitas ekuivalensi dari kedua instrumen tersebut.¹⁸

Metode bentuk paralel memiliki keunggulan kompetitif yang signifikan, antara lain mampu meminimalkan risiko terjadinya efek belajar (*practice effect*) atau efek ingatan yang sering kali mengontaminasi keaslian skor pada metode tes ulang (*test-*

¹⁶David L. Streiner, Geoffrey R. Norman, & John Cairney, *Health Measurement Scales: A Practical Guide to Their Development and Use*, 5th ed. (Oxford: Oxford University Press, 2019), 167–172.

¹⁷Saifuddin Azwar, *Reliabilitas dan Validitas* (Yogyakarta: Pustaka Pelajar, 2022), 48–52.

¹⁸Robert M. Kaplan & Dennis P. Saccuzzo, *Psychological Testing: Principles, Applications, and Issues*, 9th ed. (Boston: Cengage Learning, 2018), 130–135.

retest method). Namun, kelemahan utama metode ini terletak pada tingginya tingkat kesulitan teknis, serta besarnya alokasi waktu, tenaga, dan biaya yang dibutuhkan untuk mengonstruksi dua perangkat tes yang benar-benar ekuivalen secara statistik.¹⁹

Sebagai contoh implementasi, seorang guru PAI menyusun dua paket soal tentang materi zakat untuk dua kelompok kelas yang berbeda. Paket A dan Paket B dirancang dengan jumlah soal, distribusi tingkat kesulitan, dan cakupan materi yang setara. Setelah diberikan kepada dua kelompok yang setara, hasilnya dikorelasikan dan diperoleh nilai 0,78. Nilai ini menunjukkan bahwa kedua paket soal tersebut memiliki tingkat kesetaraan yang baik dan dapat digunakan secara bergantian tanpa mengurangi keadilan penilaian. Penelitian mengenai validitas dan reliabilitas instrumen evaluasi PAI menunjukkan bahwa guru yang secara konsisten menerapkan prosedur pengujian kualitas instrumen mampu menghasilkan data penilaian yang lebih akurat dan dapat dipertanggungjawabkan.²⁰

يَا أَيُّهَا الَّذِينَ آمَنُوا كُونُوا قَوَّامِينَ لِلَّهِ شُهَدَاءَ بِالْقِسْطِ وَلَا يَجْرِمَنَّكُمْ شَنَاٰنُ قَوْمٍ عَلَىٰ أَلَّا تَعْدِلُوا
إِعْدِلُوا هُوَ أَقْرَبُ لِلتَّقْوَىٰ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

"Wahai orang-orang yang beriman, jadilah kamu penegak (kebenaran) karena Allah (dan) saksi-saksi (yang bertindak) dengan adil. Janganlah kebencianmu terhadap suatu kaum mendorong kamu untuk berlaku tidak adil. Berlakulah adil karena (adil) itu lebih dekat pada takwa. Bertakwalah kepada Allah. Sesungguhnya Allah Mahateliti terhadap apa yang kamu kerjakan." (QS. Al-Ma'idah: 8)

Ayat ini memberikan instruksi moral yang tegas bahwa objektivitas dan keadilan harus berada di atas segala-galanya. Perangkat tes paralel yang tidak setara tingkat kesulitannya akan melahirkan ketidakadilan yang mencederai hak peserta didik. Menjaga kualitas dan kesetaraan instrumen adalah wujud pemenuhan amanah profesionalisme guru. Sebagaimana sabda Rasulullah saw. yang diriwayatkan oleh Imam al-Bukhari:

إِذَا صُيِّعَتِ الْأَمَانَةُ فَانْتَظِرِ السَّاعَةَ. قَالَ: كَيْفَ إِصَاعَتُهَا؟ قَالَ: إِذَا وُسِّدَ الْأَمْرُ إِلَىٰ غَيْرِ أَهْلِهِ
فَانْتَظِرِ السَّاعَةَ

"Apabila amanah sudah disia-siakan, maka tunggulah terjadinya kiamat." Orang itu bertanya, "Bagaimana menyia-nyiakannya?" Nabi saw. menjawab, "Apabila suatu urusan diserahkan bukan kepada ahlinya, maka tunggulah terjadinya kiamat." (HR. Al-Bukhari)²¹

Hadis ini memperingatkan dengan keras bahwa penyerahan suatu tugas kepada individu yang tidak memiliki kompetensi akan mendatangkan kerusakan. Penyusunan

¹⁹Saifuddin Azwar, *Reliabilitas dan Validitas* (Yogyakarta: Pustaka Pelajar, 2022), 55–60.

²⁰Fitri Wulandari dan Achmad Sul-toni, "Analisis Reliabilitas Instrumen Penilaian Autentik pada Mata Pelajaran PAI di MTs," *Jurnal Pendidikan Islam* 9, no. 2 (2023): 115–130; lihat juga Putu Gde Subhaktiyasa, "Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif: Sebuah Studi Pustaka," *Journal of Education Research* 5, no. 4 (2024): 5599–5609. <https://doi.org/10.37985/jer.v5i4.1138>.

²¹Hadis riwayat Muhammad al-Bukhari dalam *Sahih al-Bukhari*, Kitab al-'Ilm.

instrumen evaluasi yang valid, reliabel, dan ekuivalen memerlukan keahlian teknis psikometris yang memadai. Guru PAI selaku pemegang amanah edukatif wajib mendalami ilmu evaluasi dan statistika pendidikan agar mampu menguji instrumen penilaiannya secara ilmiah.

E. Komparasi Metode Belah Dua dan Bentuk Paralel

Setelah mengkaji masing-masing metode secara mendalam, perlu dilakukan komparasi naratif untuk memahami kapan dan bagaimana setiap metode paling tepat diterapkan dalam evaluasi pembelajaran PAI. Metode belah dua unggul dalam hal kepraktisan dan efisiensi: hanya dibutuhkan satu perangkat tes dan satu sesi pengujian, sehingga sangat ideal untuk situasi di mana sumber daya waktu dan anggaran terbatas. Akan tetapi, validitas hasilnya sangat bergantung pada cara pembelahan soal pembelahan yang tidak tepat dapat menghasilkan koefisien yang menyesatkan.

Sebaliknya, metode bentuk paralel menawarkan jaminan keadilan yang lebih robust karena mengeliminasi efek belajar dan efek ingatan. Metode ini sangat disarankan untuk konteks penilaian sumatif berskala besar, ujian susulan, atau situasi di mana integritas tes menjadi prioritas utama. Tantangan terbesar metode ini adalah intensitas sumber daya yang tinggi dalam proses konstruksi dua paket tes yang benar-benar ekuivalen. Secara ringkas, perbandingan karakteristik kedua metode tersebut disajikan dalam tabel berikut:

Tabel A.1 Perbandingan Karakteristik Metode Pengujian Reliabilitas

No	Aspek Komparasi	Metode Belah Dua (Split-Half)	Metode Bentuk Paralel (Equivalence)
1	Jumlah Perangkat Tes	1 Perangkat Tes	2 Perangkat Tes (Ekuivalen)
2	Jumlah Sesi Pengujian	1 Kali Pengujian	2 Kali Pengujian (Berurutan)
3	Fokus Konsistensi	Konsistensi Internal Butir Soal	Konsistensi Lintas-Bentuk Instrumen
4	Formula Koreksi Utama	Korelasi Pearson & Spearman-Brown	Korelasi Product Moment Pearson
5	Kelebihan Utama	Praktis, Efisien waktu & biaya	Bebas efek belajar, ideal untuk susulan
6	Kekurangan Utama	Tergantung pada cara pembelahan soal	Sangat sulit menyusun soal benar-benar setara

Sumber: Data diolah dari Azwar (2022) dan Kaplan & Saccuzzo (2018).

KESIMPULAN

Kajian konseptual ini telah membuktikan bahwa reliabilitas tes merupakan parameter psikometris yang fundamental dalam penyusunan instrumen evaluasi pembelajaran PAI. Melalui analisis komparatif yang sistematis, artikel ini menegaskan bahwa metode belah dua dan metode bentuk paralel menawarkan pendekatan metodologis yang berbeda namun saling melengkapi: metode belah dua unggul dalam efisiensi dan kepraktisan pengujian konsistensi internal, sedangkan metode bentuk paralel menjamin keadilan lintas-instrumen yang lebih robust meski menuntut investasi sumber daya yang lebih besar. Kontribusi teoretis khas artikel ini yang membedakannya dari kajian-kajian reliabilitas sebelumnya adalah sintesis integratif antara kerangka psikometris-statistik kedua metode tersebut dengan landasan aksiologis-teologis pendidikan Islam: prinsip keadilan ('adalah), ketelitian mendalam (itqan), tanggung jawab (mas'uliyah), dan profesionalisme yang bersumber dari Al-Qur'an dan Sunnah. Dengan demikian, penerapan metode reliabilitas bukan sekadar pemenuhan regulasi akademik, melainkan juga pengejawantahan nilai-nilai Islam dalam praktik evaluasi pendidikan.

Perlu dicatat bahwa kajian ini bersifat konseptual berbasis studi kepustakaan dan belum diuji secara empiris pada instrumen evaluasi PAI di lapangan. Keterbatasan ini membuka ruang yang legitimate bagi penelitian lanjutan untuk memvalidasi temuan-temuan konseptual artikel ini melalui uji empiris, sehingga relevansinya bagi praktik evaluasi pendidikan keagamaan dapat dibuktikan secara lebih kuat.

Saran

1. **Bagi Pendidik:** Para guru, khususnya guru Pendidikan Agama Islam, diharapkan terus meningkatkan kompetensi pedagogis dan profesional dalam bidang evaluasi pembelajaran, terutama dalam penguasaan teknik-teknik statistik pengujian instrumen seperti metode belah dua dan bentuk paralel. Penggunaan aplikasi manajer referensi seperti Zotero atau Mendeley juga sangat direkomendasikan untuk mendukung ketertiban pengelolaan sumber ilmiah.
2. **Bagi Lembaga Pendidikan:** Pihak sekolah, madrasah, maupun perguruan tinggi disarankan untuk memfasilitasi dan menyelenggarakan program pelatihan berkala mengenai pengembangan dan penjaminan mutu instrumen evaluasi bagi para pendidik, sehingga kapasitas teknis pengujian reliabilitas dapat ditingkatkan secara sistemik.
3. **Bagi Peneliti Selanjutnya:** Mengingat keterbatasan konseptual kajian ini, para peneliti masa depan diharapkan dapat memperluas cakupan kajian dengan melakukan penelitian eksperimental empiris yang mengomparasikan berbagai metode reliabilitas seperti Alpha Cronbach, Kuder-Richardson, dan estimasi Teori Respons Butir pada instrumen evaluasi PAI di lapangan, serta

mengeksplorasi lebih lanjut integrasi nilai-nilai Islam dalam kerangka pengembangan instrumen evaluasi keagamaan.

REFERENSI

- Al-Bukhari, Muhammad bin Ismail. Sahih al-Bukhari. Beirut: Dar al-Fikr, tt.
- Al-Qur'an al-Karim. Terjemahan Kementerian Agama RI. Jakarta: Lajnah Pentashihan Mushaf Al-Qur'an, 2019.
- Arifin, Zainal. Evaluasi Pembelajaran. Bandung: Remaja Rosdakarya, 2021.
- Azwar, Saifuddin. Reliabilitas dan Validitas. Edisi ke-4. Yogyakarta: Pustaka Pelajar, 2022.
- Creswell, John W., dan J. David Creswell. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 5th ed. Los Angeles: Sage Publications, 2018.
- Fernando, Derri Adi, dan Fajri Ismail. "Pentingnya Validitas dan Reliabilitas Instrument Evaluasi pada Pembelajaran Pendidikan Agama Islam." *TARBAWI: Jurnal Pendidikan dan Keagamaan* 11, no. 01 (2023)
- Hakim, Riza Aulia, Itsar Bolo Rangka, dan Yusi Riksa Yustiana. "Validitas dan Reliabilitas Angket Motivasi Berprestasi." *FOKUS: Jurnal Bimbingan dan Konseling dalam Pendidikan* 4, no. 4 (2021): 308–317. <https://doi.org/10.22460/fokus.v4i4.8282>.
- Kaplan, Robert M., dan Dennis P. Saccuzzo. Psychological Testing: Principles, Applications, and Issues. 9th ed. Boston: Cengage Learning, 2018.
- Mertler, Craig A. Introduction to Educational Research. 2nd ed. Los Angeles: SAGE, 2020.
- Moleong, Lexy J. Metodologi Penelitian Kualitatif. Ed. revisi. Bandung: PT Remaja Rosdakarya, 2021.
- Muharromah, Fityatul, Farid Al Fani, dan Ahmad Rosyid Ridho. "Reliabilitas Alat Ukur: Jenis-Jenis, Cara Pengukuran, dan Faktor-Faktor yang Memengaruhinya dalam Pembelajaran." *Jurnal Edukasi* 13, no. 01 (2025)
- Nunnally, Jum C., dan Ira H. Bernstein. Psychometric Theory. 3rd ed. New York: McGraw-Hill, 1994.
- Purwanto, Ngalm. Prinsip-Prinsip dan Teknik Evaluasi Pengajaran. Bandung: Remaja Rosdakarya, 2019.
- Retnawati, Heri. Analisis Kuantitatif Instrumen Penelitian. Yogyakarta: Parama Publishing, 2016.
- Setyaedhi, Hari Sugiharto. "Comparative Test of Cronbach's Alpha Reliability Coefficient, KR-20, KR-21, and Split-Half Method." *Journal of Education Research and Evaluation* 8, no. 1 (2024): 47–57. <https://doi.org/10.23887/jere.v8i1.68164>.

- Streiner, David L., Geoffrey R. Norman, dan John Cairney. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 5th ed. Oxford: Oxford University Press, 2019.
- Subhaktiyasa, Putu Gde. "Evaluasi Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif: Sebuah Studi Pustaka." *Journal of Education Research* 5, no. 4 (2024): 5599–5609. <https://doi.org/10.37985/jer.v5i4.1138>.
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*. Jakarta: RajaGrafindo Persada, 2020.
- Sugiyono. *Metode Penelitian Kualitatif*. Bandung: Alfabeta, 2018.
- Sugiyono. *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, 2022.
- Suprananto, dan Kusaeri. *Pengukuran dan Penilaian Pendidikan*. Yogyakarta: Graha Ilmu, 2012.
- Tan, Oon-Seng, ed. *Problem-Based Learning Innovation: Using Problems to Power Learning in the 21st Century*. Singapore: Cengage Learning, 2019.
- UIN Maulana Malik Ibrahim, dan Vick Ainun Haq. "Menguji Validitas dan Reliabilitas pada Mata Pelajaran Al Qur'an Hadits Menggunakan Korelasi Produk Moment-Spearman Brown." *An-Nawa: Jurnal Studi Islam* 4, no. 1 (2022)
- Wahyudi, Imam. *Pengembangan Pendidikan: Strategi Inovatif dan Kreatif dalam Mengelola Pendidikan Secara Komprehensif*. Jakarta: Prestasi Pustaka, 2018.
- Widoyoko, Eko Putro. *Evaluasi Program Pembelajaran: Panduan Praktis bagi Pendidik dan Calon Pendidik*. Yogyakarta: Pustaka Pelajar, 2020.
- Wulandari, Fitri, dan Achmad Sultoni. "Analisis Reliabilitas Instrumen Penilaian Autentik pada Mata Pelajaran PAI di MTs." *Jurnal Pendidikan Islam* 9, no. 2 (2023)
- Zed, Mestika. *Metode Penelitian Kepustakaan*. Jakarta: Yayasan Obor Indonesia, 2008.
- Zulkifli. "Pengembangan Instrumen Evaluasi Berbasis Nilai-Nilai Islami pada Pembelajaran Pendidikan Agama Islam." *Ta'dib: Jurnal Pendidikan Islam* 24, no. 1 (2019)