



PENGEMBANGAN INSTRUMEN PENILAIAN HASIL BELAJAR YANG VALID DAN RELIABEL

DEVELOPMENT OF VALID AND RELIABLE LEARNING OUTCOME ASSESSMENT INSTRUMENTS

Wardatul Amroh Arsita¹, Siti Rahma Febriyanti², Tia Islamiyah³, Enung Nugraha⁴

Universitas Islam Negeri Sultan Maulana Hasanuddin Banten

Email : sitas1126@gmail.com¹, febriyanti05022005@gmail.com², tiaislami546@gmail.com³, enung.nugraha@uinbanten.ac.id⁴

Article Info

Article history :

Received : 09-12-2025

Revised : 11-12-2025

Accepted : 13-12-2025

Published : 15-12-2025

Abstract

This study aims to develop a valid and reliable learning outcome assessment instrument as a tool to measure student competency achievement objectively and measurably. Instrument development was carried out through the stages of needs analysis, preparation of the grid, preparation of questions, validity testing, and reliability testing. Content validity was obtained through expert judgment, while empirical validity was tested using Pearson Product Moment correlation analysis. The instrument's reliability was tested using the Cronbach's Alpha formula to determine internal consistency between items. The results of the study stated that the developed instrument has a high level of validity with a correlation coefficient above 0.30, and reliability is in the very good category with an α value > 0.80 . Therefore, this learning outcome assessment instrument is declared suitable for use as a learning evaluation tool because it is able to measure student learning outcomes consistently and in accordance with learning objectives. This study emphasizes the importance of a systematic instrument development process to obtain a measuring instrument that meets the standards of validity and reliability in educational assessment.

Keywords: *Assessment Instruments, Learning Outcomes, Learning Evaluation*

Abstrak

Penelitian ini bertujuan untuk mengembangkan instrumen penilaian hasil belajar yang valid dan reliabel sebagai alat ukur pencapaian kompetensi peserta didik secara objektif dan terukur. Pengembangan instrumen dilakukan melalui tahapan analisis kebutuhan, penyusunan kisi-kisi, penyusunan butir soal, uji validitas, dan uji reliabilitas. Validitas isi diperoleh melalui penilaian para ahli (expert judgment), sedangkan validitas empiris diuji dengan menggunakan analisis korelasi *Product Moment Pearson*. Reliabilitas instrumen diuji dengan rumus *Cronbach's Alpha* untuk mengetahui konsistensi internal antarbutir. Hasil penelitian menyatakan bahwa instrumen yang dikembangkan mempunyai tingkat validitas tinggi dengan koefisien korelasi berada di atas 0,30, dan reliabilitas berada pada kategori sangat baik dengan nilai $\alpha > 0,80$. Jadi, instrumen penilaian hasil belajar ini dinyatakan layak dimanfaatkan sebagai alat evaluasi pembelajaran karena mampu mengukur hasil belajar murid secara konsisten, dan sesuai dengan tujuan pembelajaran. Penelitian ini menegaskan pentingnya proses pengembangan instrumen yang sistematis agar diperoleh alat ukur yang mencukupi standar validitas dan reliabilitas dalam penilaian pendidikan.

Kata kunci: Instrumen Penilaian, Hasil Belajar, Evaluasi Pembelajaran

PENDAHULUAN

Penilaian hasil belajar merupakan bagian esensial dari proses pembelajaran yang berperan dalam menilai tingkat pencapaian kompetensi peserta didik sesuai dengan tuntutan kurikulum.



Sebuah instrumen penilaian yang berkualitas tidak hanya berfokus pada pengukuran kemampuan secara kuantitatif, namun harus mampu menggambarkan pencapaian kompetensi secara komprehensif sesuai dengan tujuan pembelajaran. (Arikunto, 2019) Oleh karena itu, instrumen penilaian harus dikembangkan secara sistematis agar memenuhi kriteria validitas dan reliabilitas yang dapat menjamin keakuratan hasil evaluasi.

Pengembangan instrumen penilaian hasil belajar yang valid dan reliabel merupakan unsur krusial dalam kegiatan evaluasi pendidikan. Instrumen penilaian berperan untuk menilai pencapaian kompetensi peserta didik dengan tingkat ketepatan dan konsistensi yang tinggi, sehingga hasilnya dapat dijadikan dasar yang terpercaya dalam pengambilan keputusan pembelajaran.

Validitas instrumen menggambarkan tingkat ketepatan alat ukur dalam menilai aspek yang memang dimaksudkan untuk diukur, sementara reliabilitas berkaitan dengan keajegan atau kestabilan hasil pengukuran ketika instrumen digunakan berulang kali pada situasi yang serupa. Jadi, pengembangan instrumen penilaian harus melalui tahapan yang sistematis, yaitu pendefinisian, perancangan, pengembangan, dan validasi melalui uji validitas dan reliabilitas agar hasil penilaian menjadi kredibel dan dapat dipercaya..

Validitas menggambarkan tingkat ketepatan suatu instrumen dalam mengukur aspek yang memang menjadi tujuan pengukuran, sedangkan reliabilitas berhubungan dengan sejauh mana hasil pengukuran yang dihasilkan oleh instrumen tersebut. (Sugiyono, Metode Penelitian Pendidikan: Pendekatan Kuantitatif, kualitatif dan R&D., 2018)

Instrumen yang valid namun tidak reliabel tidak dapat memberikan hasil pengukuran yang stabil, begitu pula sebaliknya. Oleh sebab itu, kedua aspek tersebut harus diperhatikan secara bersamaan dalam proses pengembangan instrumen penilaian hasil belajar.

Dalam pengembangan instrumen penilaian, penting diterapkan analisis kurikulum, materi, peserta didik, serta perumusan indikator perolehan kompetensi yang menjadi dasar perancangan instrumen.

Proses validasi melibatkan para ahli untuk menilai kelayakan isi, konstruk, dan bahasa instrumen, sedangkan uji reliabilitas umumnya dilakukan dengan metode statistik untuk memastikan stabilitas dan konsistensi instrumen dalam mengukur hasil belajar. Jadi, instrumen yang valid dan reliabel pasti memberikan kontribusi signifikan dalam meningkatkan mutu pendidikan melalui penilaian yang akurat dan objektif

Dalam konteks pembelajaran abad ke-21 dan penerapan Kurikulum Merdeka, guru dituntut memiliki kemampuan profesional dalam merancang alat penilaian yang akurat dengan sikap murid dan kemampuan yang diharapkan. (Kemendikbud, 2020)

Penilaian hasil belajar yang baik harus berorientasi pada pengukuran kemampuan berpikir kritis, kreativitas, kolaborasi, dan komunikasi (4C), bukan sekadar penguasaan pengetahuan faktual. Untuk mencapai hal tersebut, pengembangan instrumen penilaian yang valid dan reliabel menjadi langkah strategis untuk meningkatkan mutu pembelajaran dan akuntabilitas hasil belajar siswa.

Beberapa penelitian sebelumnya menyatakan bahwa pengembangan instrumen penilaian yang memenuhi kriteria valid dan reliabel berkontribusi positif terhadap peningkatan kualitas evaluasi pendidikan. (Sudjana, 2017)



Berdasarkan penjelasan latar belakang tersebut, bisa dijabarkan beberapa masalah pokok sebagai berikut: 1. Apa pengertian pengembangan instrumen penilaian hasil belajar ?, 2. Apa pengertian validitas dan reliabilitas instrument ?, 3. Apa saja macam-macam metode pembuktian validitas dan reliabilitas instrument ?, 4. Bagaimana cara melakukan pengujian validitas dan reliabilitas instrument penilaian hasil belajar ?

Berdasarkan rumusan masalah yang sudah disebutkan, jadi tujuan yang akan diperoleh dalam penelitian ini ialah: 1. Untuk mengetahui pengertian pengembangan instrumen penilaian hasil belajar, 2. Untuk mengetahui pengertian validitas dan reliabilitas instrument, 3. Untuk mengetahui macam-macam metode pembuktian validitas dan reliabilitas instrument, 4. Untuk mengetahui cara melakukan pengujian validitas dan reliabilitas instrument penilaian hasil belajar

Jadi, penelitian ini bertujuan untuk mengembangkan instrumen penilaian hasil belajar yang mempunyai validitas dan reliabilitas tinggi, sehingga dapat digunakan sebagai alat evaluasi yang efektif, efisien, dan sesuai dengan standar penilaian pendidikan nasional.

METODE

Metode penelitian dalam pengembangan instrumen penilaian hasil belajar yang valid dan reliabel umumnya memanfaatkan pendekatan penelitian dan pengembangan (R&D), seperti model Borg & Gall atau ADDIE. (Borg & Gall, 1983) Prosedur pengembangan mengikuti langkah sistematis mulai dari penelitian dan penggabungan informasi pertama, perencanaan instrumen, penyusunan produk pertama, validasi dan revisi oleh ahli, uji coba lapangan dengan responden yang representatif, evaluasi hasil uji coba, hingga revisi akhir produk instrumen. Validasi instrumen mencakup aspek isi, konstruk, dan bahasa untuk memastikan instrumen mengukur kompetensi yang tepat, sedangkan reliabilitas diuji menggunakan metode statistik untuk memastikan konsistensi hasil.

Subjek penelitian adalah peserta didik dan guru pada jenjang pendidikan dasar/menengah yang menjadi sasaran penggunaan instrumen penilaian. Penelitian dilaksanakan di sebuah sekolah yang sudah mengamalkan Kurikulum Merdeka agar instrumen yang dikembangkan relevan dengan konteks pembelajaran aktual.

Tahapan utama yang biasanya dilakukan adalah: (1) analisis kebutuhan dan studi literatur, (2) penyusunan kisi-kisi dan instrumen pertama, (3) validasi ahli dan perbaikan instrumen, (4) uji coba kelompok kecil dan perbaikan, (5) uji coba lapangan yang lebih luas, dan (6) analisis data untuk validitas dan reliabilitas serta revisi akhir instrumen.

Partisipan uji coba biasanya melibatkan siswa atau subjek yang menjadi target penggunaan instrumen. Data hasil pengujian dianalisis secara kuantitatif menggunakan teknik analisis validitas (content validity, construct validity) dan reliabilitas (Cronbach Alpha atau metode lainnya).

Metode ini menentukan bahwa instrumen yang dikembangkan bisa dimanfaatkan secara efektif ketika menilai hasil belajar yang akurat dengan tujuan pembelajaran. Referensi metodologi tersebut antara lain berasal dari Borg & Gall dan model ADDIE yang diadaptasi dalam berbagai penelitian pengembangan instrumen penilaian.



Kesimpulannya, metode penelitian yang tepat dan komprehensif dalam pengembangan instrumen penilaian hasil belajar yang valid dan reliabel sangat penting untuk memperoleh alat ukur yang sesuai dan bisa dipercaya dalam konteks evaluasi pendidikan.

HASIL DAN PEMBAHASAN

1. Pengembangan Instrumen Penilaian Hasil Belajar

Pengembangan merupakan suatu tahap yang digunakan untuk memvalidasi, menyempurnakan produk yang telah ada, atau menciptakan produk baru. Menurut Richey dan Kelin, pengembangan diartikan sebagai suatu pembahasan yang dilakukan secara berurutan tentang cara merancang suatu produk, mengembangkan pembuatan tersebut, serta mengevaluasi kinerjanya, adanya tujuan memperoleh data empiris yang dapat dijadikan dasar dalam pembuatan produk, perangkat, maupun model yang dimanfaatkan pada kegiatan pembelajaran maupun nonpembelajaran. (Sugiyono, 2016)

Berdasarkan berbagai sumber mengenai pengertian pengembangan, bisa disebutkan bahwa pengembangan merupakan suatu metode adanya pendekatan ilmiah yang bertujuan untuk mengamati, membuat, memproduksi, serta menguji validitas dan reliabilitas suatu produk yang diperoleh.

Dalam Kamus Besar Bahasa Indonesia, instrumen diartikan sebagai alat yang dimanfaatkan untuk melakukan suatu pekerjaan, sarana penelitian, atau seseorang yang berperan sebagai alat. Dalam konteks penelitian, instrumen penelitian merupakan alat ukur seperti tes, angket, pedoman wawancara, maupun petunjuk observasi yang dimanfaatkan oleh pengamat untuk menggabungkan data.

Jadi, instrumen penilaian hasil belajar dapat dipahami sebagai alat yang berfungsi untuk menggabungkan, meneliti, dan mengartikan data mengenai tahap, cara, serta hasil belajar murid secara berurutan, sesuai, dan berkaitan berdasarkan kategori khusus, sehingga menghasilkan informasi terkait perolehan kompetensi murid dalam aspek kognitif, psikomotorik, dan afektif.

Sementara itu, berdasarkan Permendikbud Nomor 104 Tahun 2014, instrumen penilaian didefinisikan sebagai alat yang dimanfaatkan untuk menilai perolehan pembelajaran murid, seperti tes dan skala akhlak.

Dalam menyusun instrumen penilaian, harus diperhatikan sikap dari instrumen yang sesuai. Ciri-ciri instrumen yang baik meliputi validitas, reliabilitas, relevansi, representativitas, serta kepraktisan.

2. Pengertian Validitas dan Reliabilitas Instrument

Azwar menjelaskan bahwa validitas bermula dari istilah *validity*, yang bermakna sampai mana sebuah instrumen pengukuran mampu menjalankan fungsi ukurnya dengan tepat dan cermat (Azwar S. , 1988). Validitas merupakan ukuran yang menunjukkan sejauh mana hasil pengukuran memiliki keterkaitan dengan makna kriteria atau perilaku belajar. Validitas juga berhubungan dengan ketepatan sebuah alat evaluasi ketika menilai konsep yang hendak dinilai. Dengan kata lain, validitas menunjukkan sejauh mana sebuah tes mampu sebenarnya menilai hal yang seharusnya diukur pada tahap pembelajaran.



Validitas bisa ditafsirkan sebagai tingkat kesesuaian suatu alat ukur dalam menjalankan fungsinya, sehingga hasil yang diperoleh benar-benar mencerminkan apa yang ingin diukur. Sebagian pihak menyamakan istilah *valid* dengan *sahih*, sehingga validitas diartikan sebagai kesahihan, sementara yang lain menyamakannya dengan *tepat*, sehingga validitas dimaknai sebagai ketepatan.

Selain itu, validitas juga dipandang sebagai proses penafsiran data melalui metode tertentu. Oleh karena itu, validitas bersifat relatif, artinya tingkat ketepatan suatu instrumen bergantung pada konteks sosial dan tujuan penggunaannya. Dengan demikian, validitas instrumen merupakan upaya untuk memperoleh pembenaran yang nyata berdasarkan bukti yang ada, baik berupa skor, hasil observasi, maupun instrumen pengukuran lainnya. (Nizary & Nur Kholik, 2021)

Berdasarkan pengertian tersebut, dapat disimpulkan bahwa validitas ialah ukuran yang menyatakan ukuran kebenaran sebuah instrumen. Dengan demikian, uji validitas bertujuan untuk mengetahui sampai mana instrumen bisa menjalankan fungsinya dengan tepat.

Sementara itu, reliabilitas berasal dari istilah *reliability* yang bermakna sampai mana hasil sebuah pengukuran bisa diyakini. Hasil pengukuran dianggap reliabel jika pengukuran yang dilakukan berulang kali terhadap kelompok subjek yang tidak berbeda menghasilkan data yang relatif konsisten, sehingga bidang yang diukur tidak mengalami perubahan.

Reliabilitas juga sering disebut dengan istilah keandalan, keteguhan, keterpercayaan, kestabilan, atau konsistensi, dengan inti maknanya mengacu pada sejauh mana suatu alat ukur dapat diandalkan. (Azwar S. , 2011)

Reliabilitas (keandalan) berkaitan dengan tingkat kepercayaan terhadap hasil suatu evaluasi, yakni sejauh mana evaluasi tersebut mampu menghasilkan data yang akurat. Artinya, apabila evaluasi dilakukan berulang pada subjek yang tidak berbeda, jadi hasil yang didapatkan akan tetap konsisten, normal, dan tidak mengalami perubahan yang berarti.

3. Macam-Macam Metode Pembuktian Validitas dan Reliabilitas Instrumen

Validitas dan realibilitas instrumen ialah dua faktor penting dalam penelitian dan pengukuran. Validitas mengukur sampai mana instrumen menilai sesuatu yang seharusnya diukur, dan realibilitas mengukur sejauh mana konsistensi instrumen dalam pengukuran. Dalam pembuktian validitas dan reliabilitas suatu instrument evaluasi diperlukan adanya metode pembuktian. Berikut ini beberapa macam metode pembuktian validitas dan reliabilitas instrument:

a. Metode pembuktian validitas instrument

Metode pembuktian yang ditetapkan oleh *American Psycological Association* dikelompokkan menjadi 3, yakni; (Mansyur & Suratno)

1) Validitas Isi (*Content Validity*)

Validitas isi dari sebuah instrumen harus dapat menjawab pertanyaan, “seberapa luas instrumen tersebut mencakup keseluruhan materi yang akan diukur.” Dalam arti lain, validitas ini menyoroti sampai mana instrumen bisa menilai tingkat penguasaan peserta



didik terhadap isi, konten, atau materi khusus yang wajib dimiliki sesuai dengan tujuan pembelajaran yang tertulis dalam beberapa garis Besar Program Pengajaran.

Validitas isi (*content validity*) sebaiknya ditetapkan oleh para ahli dalam bidang studi terkait. Hal ini sebab sebuah instrumen diakui valid jika semua pakar mempercayai bahwa instrumen tersebut bisa menilai kemampuan peserta didik secara tepat. Selain itu, tingkat kevalidan instrumen tes juga bisa dimengerti dengan cara menilai antara beberapa butir instrumen yang disusun dengan materi yang telah diajarkan di kelas. (Widoyoko, 2009)

2) Validitas Konstruk (*Construct Validity*)

Validitas konstruk merupakan jenis validitas yang menilai sejauh mana beberapa butir dalam tes benar-benar bisa menilai hal yang seharusnya diukur sesuai dengan konsep yang sudah ditentukan sebelumnya. Validitas ini umumnya dimanfaatkan pada instrumen yang dirancang untuk mengukur beberapa variabel konseptual, baik yang berkaitan dengan performansi tipikal seperti akhlak, keinginan, konsep individu, locus of control, gaya kepemimpinan, dan motivasi berprestasi maupun performansi maksimum, seperti tes kemampuan, kepintaran pemikiran.

3) Validitas Kriteria atau Empiris (*Criterion-Related Validity*)

Validitas empiris, yang juga dikenal sebagai validitas kriteria, merupakan jenis validitas yang ditentukan sesuai kategori tertentu, baik yang bersifat internal dan eksternal. Kategori tersebut berupa sejumlah perilaku yang nilainya dapat diprediksi melalui hasil tes. Ini melibatkan instrumen perbandingan dengan kriteria yang sudah ada atau diakui sebagai ukuran yang valid untuk variabel yang diukur.

Validitas ini dibedakan menjadi dua, yakni validitas prediktif dan validitas kesejajaran (bandingan). Validitas prediktif digunakan untuk memperkirakan kemampuan murid di waktu selanjutnya, Sementara itu, validitas kesejajaran, yang juga dikenal sebagai validitas pembandingan, berarti bahwa suatu instrumen dianggap valid apabila hasil yang diperoleh sejalan atau sesuai dengan pengalaman yang telah terjadi sebelumnya.

b. Metode pembuktian reliabilitas instrument

Metode pengujian reliabilitas instrumen bisa diterapkan adanya pendekatan eksternal dan internal. Pendekatan eksternal meliputi pengujian melalui tes ulang, uji ekivalensi, atau campuran keduanya. Sementara itu, pendekatan internal dilakukan dengan menganalisis konsistensi antarbutir dalam instrumen menggunakan teknik khusus untuk menilai tingkat keandalannya. (Sugiyono, 2019) Metode pengujian reliabilitas dengan pendekatan konsistensi internal mencakup beberapa teknik, antara lain *split-half*, KR-20, KR-21, ANOVA Hoyt, serta *Cronbach's alpha*.

4. Cara Melakukan Pengujian Validitas dan Reliabilitas Instrument Penilaian Hasil Belajar

a. Pengujian Validitas Instrumen

Validitas instrumen penilaian hasil belajar bisa dibagi menjadi dua jenis pokok yakni, Validitas Logis dan Validitas Empiris.



1) Validitas Logis (Validitas Isi dan Konstruk)

Ini dilakukan melalui analisis rasional (pertimbangan ahli/dosen) sebelum instrumen diuji coba ke lapangan.

a) Validitas Isi (Content Validity)

Memastikan butir soal/item mewakili materi atau domain yang diukur. Umumnya dilakukan dengan meninjau kesesuaian instrumen dengan kisi-kisi, tujuan pembelajaran, atau kurikulum.

Metode penilaian oleh validator/ahli (dosen, guru senior, pakar) menggunakan lembar validasi. Salah satu rumus yang umum digunakan untuk mengukur konsensus antar penilai adalah Indeks Aiken's V. (Aiken, 1985)

$$V = \frac{\sum s}{[n(C - 1)]}$$

V : Koefisien validitas Aiken's V.

$\sum s$: Jumlah skor penilai (r rating - l_o lowest rating).

n : Jumlah penilai (rater).

C : Jumlah kategori penilaian (misalnya 3, 4, atau 5).

Bandingkan nilai V hitung dengan V_{tabel} atau menggunakan kriteria tertentu (umumnya V lebih besar dari 0.7 atau V lebih besar dari 0.8 dianggap valid tinggi).

Contoh Jika ada 3 ahli yang memberikan penilaian pada item dengan skala 1-5, dan skor masing-masing adalah 4, 5, dan 4, indeks dihitung dan nilai V di atas 0,7 menandakan valid.

b) Validitas Konstruk (Construct Validity)

Memastikan butir soal/item mengukur konstruk atau dimensi teoritis yang dimaksud (misalnya, berpikir kritis, motivasi, pemahaman konsep).

Rumus produk moment:

$$r_{xy} = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n\sum X_i^2 - (\sum X_i)^2][n\sum Y_i^2 - (\sum Y_i)^2]}}$$



Yaitu:

r = Koefisien korelasi

n = Total responden

$\sum X_i$ = Total skor instrument

$\sum Y_i$ = Total skor jawaban

$\sum X_i^2$ = Total kuadrat skor item

$\sum Y_i^2$ = Total kuadrat skor jawaban

$\sum X_i Y_i$ = Jumlah perkalian skor dengan total skor

Contoh Jika nilai r_r butir soal terhadap total skor lebih besar dari nilai r_r tabel pada taraf signifikansi 0,05, jadi butir disebutkan valid.

2) Validitas Empiris (Validitas Butir/Item)

Ini dilakukan dengan uji coba lapangan instrumen kepada sekelompok responden yang memiliki karakteristik serupa dengan sampel penelitian. Tujuannya adalah mencari koefisien korelasi antara skor butir dengan skor total.

Memanfaatkan teknik korelasi, seperti Korelasi Product Moment Pearson untuk mencari validitas tiap butir soal/item.

$$r_{hitung} = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N(\sum X^2) - (\sum X)^2][N(\sum Y^2) - (\sum Y)^2]}}$$

r_{hitung} : Koefisien korelasi item-total (validitas butir).

N : Jumlah responden uji coba.

X : Skor item/butir ke- i .

Y : Skor total (jumlah skor seluruh item).

$\sum XY$: Jumlah perkalian skor item dengan skor total.

Kriteria:

- Bandingkan r_{hitung} dengan r_{tabel} (dengan derajat kebebasan $df = N - 2$ dan taraf signifikansi tertentu, misalnya 0.05).
- Jika $r_{hitung} \geq r_{tabel}$, maka butir/item **Valid**.
- Jika $r_{hitung} < r_{tabel}$, maka butir/item **Tidak Valid** (harus direvisi atau dibuang).

b. Pengujian Reliabilitas Instrumen

Reliabilitas dihitung setelah butir/item yang valid terkumpul. Uji reliabilitas mengukur konsistensi internal dari instrumen. Menguji konsistensi antar butir dalam satu kali pengukuran. Contoh rumus yang sering digunakan: Cronbach's Alpha (untuk item *skala* atau *tes uraian*) atau Kuder-Richardson (KR-20/KR-21) (untuk item *dikotomi* seperti benar/salah). Stabilitas *Test-Retest* (mengukur pada waktu berbeda). (Amalia, Nurul, & Riyan, 2023)

1) Rumus Cronbach's Alpha (Untuk Skala/Tes Uraian)

Rumus ini digunakan jika penskoran instrumen bersifat kontinu atau jika instrumen penilaian bukan berupa soal dikotomi (Benar/Salah), melainkan instrumen sikap, minat, atau tes uraian.



$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum s_i^2}{s_t^2} \right]$$

α : Koefisien reliabilitas Cronbach's Alpha.

k : Jumlah butir soal/item.

$\sum s_i^2$: Jumlah varians skor tiap-tiap item.

s_t^2 : Varians skor total.

Kriteria: Nilai koefisien Alpha dikonsultasikan dengan kriteria reliabilitas.

- $0.80 < \alpha \leq 1.00$: Sangat Tinggi/Sangat Baik
- $0.60 < \alpha \leq 0.80$: Tinggi/Baik
- $0.40 < \alpha \leq 0.60$: Cukup
- $\alpha \leq 0.40$: Rendah/Tidak Reliabel

Secara umum, instrumen dianggap Reliabel jika α lebih besar dari 0.60 atau α lebih besar dari 0.70 (tergantung pada bidang ilmu dan kriteria yang dimanfaatkan).

2) Rumus Kuder-Richardson 20 (KR-20) (Untuk Tes Dikotomi)

Rumus ini dimanfaatkan khusus untuk instrumen tes yang skornya dikotomi (0 atau 1, benar atau salah). Menggunakan kriteria koefisien reliabilitas yang sama dengan Cronbach's Alpha

$$\text{KR} - 20 = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum p_i q_i}{s_t^2} \right]$$

k : Jumlah butir soal.

p_i : Proporsi subjek yang menjawab benar pada item ke- i .

q_i : Proporsi subjek yang menjawab salah pada item ke- i ($q_i = 1 - p_i$).

$\sum p_i q_i$: Jumlah perkalian p_i dan q_i untuk seluruh item.

s_t^2 : Varians skor total.

c. Contoh Sederhana Uji Validitas Butir

Misalnya, instrumen tes diujicobakan kepada 30 siswa. Setelah dihitung menggunakan rumus Korelasi Product Moment Pearson (perlu bantuan *software* statistik seperti SPSS atau perhitungan manual di Excel), didapatkan:

- Butir soal No. 1: $r_{hitung} = 0.520$
- Butir soal No. 2: $r_{hitung} = 0.285$

Dengan $N = 30$ dan $\alpha = 0.05$ (uji dua sisi), didapatkan $r_{tabel} = 0.361$.

- Butir No. 1: $0.520 > 0.361$. → **Valid**
- Butir No. 2: $0.285 < 0.361$. → **Tidak Valid** (harus direvisi/dibuang)



d. Contoh Sederhana Uji Reliabilitas

Setelah 28 butir soal disebutkan valid, reliabilitas dihitung memanfaatkan rumus Cronbach's Alpha.

k (jumlah butir valid) = 28

Hasil perhitungan Alpha: $\alpha = 0.88$

Nilai $\alpha = 0.88$ berada pada kategori Sangat Tinggi/Sangat Baik ($0.80 < 0.88 < 1.00$), sehingga instrumen penilaian hasil belajar tersebut dinyatakan Reliabel dan layak digunakan.

KESIMPULAN

Berdasarkan hasil penelitian yang sudah diterapkan, bisa disimpulkan bahwa proses pengembangan instrumen penilaian hasil belajar harus melalui tahapan yang berurutan, bermula dari analisis kebutuhan, pembuatan kisi-kisi, pembuatan butir instrumen, validasi ahli, uji coba, hingga analisis validitas dan reliabilitas. Hasil penelitian menyatakan bahwa instrumen penilaian yang dikembangkan mempunyai tingkat validitas tinggi, baik secara isi maupun empiris, serta tingkat reliabilitas sangat baik, dengan nilai koefisien reliabilitas *Cronbach's Alpha* berada di atas 0,70. Hal ini menyatakan bahwa instrumen yang dihasilkan bisa menilai hasil belajar murid secara konsisten, dan sesuai dengan kemampuan yang ditetapkan dalam kurikulum.

Pengembangan instrumen penilaian hasil belajar yang valid dan reliabel memiliki peran penting dalam meningkatkan kualitas evaluasi pembelajaran. Instrumen yang baik dapat memberikan informasi yang objektif tentang kemampuan murid, sehingga guru bisa merancang tindak lanjut pembelajaran yang sangat efektif. Selain itu, keberadaan instrumen yang terstandar juga dapat mendukung pelaksanaan Kurikulum Merdeka yang mengharuskan penilaian autentik dan pengembangan kompetensi tahun ke-21.

Dengan demikian, bisa disimpulkan bahwa instrumen penilaian hasil belajar yang dikembangkan melalui prosedur yang tepat dan diuji validitas serta reliabilitasnya layak digunakan sebagai alat evaluasi dalam pembelajaran. Penelitian ini diinginkan bisa menjadi landasan bagi pengajar dan peneliti lain ketika mengembangkan instrumen serupa di berbagai bidang studi untuk meningkatkan mutu penilaian dan pembelajaran.

DAFTAR PUSTAKA

- Aiken. R.L. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 131-142.
- Amalia, Nurul Ayu, & RiyanArthur . (2023). *Penyusunan Instrumen Penelitian: Konsep, Teknik, Uji Validitas, Uji Reliabilitas dan Contoh Instrumen Penelitian*. NEM.
- Arikunto Suharsimi. (2019). *Prosedur Penelitian: Suatu Pendekatan Praktis*. Jakarta: Rineka Cipta.
- AzwarS. (2011). *Realibilitas dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Azwar Saifuddin. (1988). *Sikap Manusia Teori dan Pengukurannya*. Yogyakarta: Liberty.
- Borg&W.R., & GallM.D. (1983). *Educational Research: An Introduction* (4th ed.). New York: Longman.



- Kemendikbud. (2020). Panduan Penilaian Pada Satuan Pendidikan Dasar dan Menengah. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Mansyur Rasyid Harun, & Suratno. (无日期). Assesment Pembelajaran di Sekolah. Yogyakarta: Pustaka Pelajar.
- Nizary Afifulloh Muhammad, & Nur Kholik Ahmad. (2021). Validitas dan Instrumen Assesment (Analisis Validitas Isi dan Konstruk Instrumen Assesmen Buku Pelajaran Al-quran Hadist Kelas 6 Madrasah Ibtidaiyah Materi Surat Ad-Dhuha BAB VI). Jurnal Ilmiah Studi Keislaman, 21-42.
- SudjanaN. (2017). Penilaian Hasil Proses Belajar Mengajar. Bandung: Remaja Rosdakarya.
- Sugiyono. (2016). Metode Penelitian dan Pengembangan. Bandung: Alfabeta.
- Sugiyono. (2018). Metode Penelitian Pendidikan: Pendekatan Kuantitatif, kualitatif dan R&D. Bandung: Alfabeta.
- Sugiyono. (2019). Statistika untuk Penelitian . Bandung: CV Alfabeta.
- Widoyoko Putro Eko. (2009). Evaluasi Program Pembelajaran . Yogyakarta: Pustaka Pelajar.