

ANALISIS VALIDITAS INSTRUMEN ASESMEN PADA MATERI TOKOH-TOKOH GERAKAN PEMBARUAN ISLAM DALAM BUKU SKI KELAS 12 MA MUALLIMIN MUHAMMADIYAH YOGYAKARTA

Rahmadani Akbar¹, Muhammad Fikri Noor Fadjri², Suwadi³
^{1,2,3} Universitas Islam Negeri Sunan Kalijaga, Yogyakarta, Indonesia
Alamat e-mail : ¹ 24204011004@student.uin-suka.ac.id,
² 24204011014@student.uin-suka.ac.id, ³ suwadi@uin-suka.ac.id

ABSTRACT

This study aims to analyze the validity of assessment instruments on the topic of Islamic reform movement figures in the Islamic Cultural History (SKI) textbook for Grade 12 at MA Muallimin Muhammadiyah Yogyakarta. Previous studies highlight gaps in teachers' design of assessment tools, particularly in developing questions that measure higher-order thinking skills and reflective values in Islamic history learning. Using a descriptive-analytical qualitative approach, the research focuses on content validity and construct validity analyses of five essay questions in Chapter V of the SKI textbook. Primary data sources include the Grade 12 SKI textbook and its accompanying assessment instruments, while secondary data comprise curriculum documents and related references. The findings reveal that the assessment instruments demonstrate moderate to high content validity, but their construct validity varies significantly. Most questions target low to intermediate cognitive levels (C2-C3), with only one question reaching a higher cognitive level (C4-C5). This indicates a need to refine assessment tools to not only evaluate factual knowledge but also foster critical thinking, analytical skills, and students' ability to integrate Islamic reform values into contemporary contexts.
Keywords: *Instrument Validity, Learning Assessment, Islamic Cultural History, Islamic Reformers.*

ABSTRAK

Penelitian ini bertujuan untuk menganalisis validitas instrumen asesmen pada materi tokoh-tokoh gerakan pembaruan Islam dalam buku Sejarah Kebudayaan Islam (SKI) kelas 12 MA Muallimin Muhammadiyah Yogyakarta. Berbagai studi menunjukkan adanya kesenjangan dalam penyusunan instrumen penilaian oleh guru, terutama dalam pengembangan soal yang mampu mengukur kemampuan berpikir tingkat tinggi dan nilai reflektif dalam pembelajaran sejarah Islam. Penelitian menggunakan pendekatan kualitatif deskriptif-analitis dengan fokus pada analisis validitas isi dan validitas konstruk dari lima soal uraian pada Bab V buku SKI. Sumber data primer adalah buku ajar SKI kelas 12 dan instrumen asesmen yang digunakan, sedangkan data sekunder mencakup dokumen kurikulum dan referensi terkait. Hasil penelitian menunjukkan bahwa instrumen asesmen memiliki validitas isi yang cukup hingga sangat baik, namun validitas

konstruknya bervariasi. Mayoritas soal masih berada pada level kognitif rendah hingga menengah (C2-C3), dengan hanya satu soal yang mencapai level kognitif tinggi (C4-C5). Temuan ini mengindikasikan perlunya pengembangan instrumen asesmen yang tidak hanya mengukur pemahaman faktual, tetapi juga mendorong kemampuan berpikir kritis, analitis, dan kemampuan siswa dalam mengintegrasikan nilai-nilai pembaruan Islam dalam konteks kontemporer.

Kata Kunci: Validitas Instrumen, Asesmen Pembelajaran, Sejarah Kebudayaan Islam, Tokoh Pembaruan Islam.

A. Pendahuluan

Salah satu kemampuan yang wajib dimiliki oleh seorang guru adalah memberikan evaluasi dalam proses pembelajaran (Arifin, 2015; Magdalena dkk., 2020; Sutrisno dkk., 2022). Adapun keterampilan dasar yang perlu dikuasai guru dalam evaluasi pembelajaran adalah membuat indikator-indikator penilaian berdasarkan standar penilaian yang telah ditentukan (Supriyadi, 2011). Standar penilaian ini meliputi tujuan, manfaat, prinsip, mekanisme, prosedur, dan instrument penilaian hasil belajar peserta didik sebagai dasar dalam melakukan penilaian (Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016 Tentang Standar Penilaian Pendidikan, 2016). Prinsip-prinsip yang nantinya menjadi perhatian khusus dalam melakukan penilaian adalah mengacu pada kompetensi

standar dan indikator ketercapaian hasil belajar bersifat menyeluruh, meliputi standar kompetensi, kompetensi dasar, indikator pencapaian, ataupun aspek-aspek yang terkait dengan penilaian (kognitif, afektif dan psikomotorik), bersifat berkelanjutan, artinya perencanaan penilaian dilakukan secara terus menerus serta bertahap yang bertujuan untuk mendapatkan gambaran perkembangan siswa dan menyesuaikan dengan apa yang dipelajari di sekolah (Ratnawulan & Rusdiana, 2013).

Bagi para guru, melaksanakan penilaian bukanlah hal yang mudah. Banyak di antara mereka yang menghadapi kesulitan dalam merumuskan indikator instrumen penilaian, memilih teknik penilaian yang sesuai dengan kompetensi dasar yang diajarkan, serta mengembangkan butir-butir instrumen. Dari tiga aspek penilaian,

yaitu kognitif, afektif, dan psikomotorik, guru sering kali mengalami kesulitan dalam menilai aspek sikap (afektif) dalam waktu yang terbatas dengan menggunakan berbagai teknik penilaian. Pada dasarnya, ketepatan suatu penilaian hanya berlaku untuk tujuan tertentu saja, artinya satu instrumen penilaian belum tentu dapat digunakan secara universal. Misalnya, ketika sekolah A menggunakan instrumen yang sama dengan sekolah B, hasilnya bisa saja berbeda, begitu pula sebaliknya (Sukardi, 2015).

Sebelum melakukan penilaian, guru perlu memahami bahwa setiap siswa memiliki kemampuan dan keterampilan yang berbeda-beda. Hal ini karena tidak semua siswa mampu memahami materi dalam waktu yang sama. Oleh karena itu, kemampuan guru dalam memetakan tingkat pemahaman semua siswa sangat penting untuk memudahkan proses penilaian nantinya (Kemendikbud, 2020). Perlu diingat bahwa penilaian terdiri dari berbagai bentuk, seperti penilaian harian (PH) yang dilakukan guru untuk mengukur pencapaian kompetensi dasar. Selain itu, ada juga penilaian tengah semester dan akhir semester, yaitu kegiatan

penilaian yang dilakukan guru untuk mengukur kemampuan siswa setelah menjalani pembelajaran selama kurang lebih 3 hingga 6 bulan. Berbagai bentuk penilaian ini pada dasarnya bertujuan untuk mencapai tujuan pembelajaran, yaitu perubahan dalam diri siswa yang mencakup aspek kognitif, afektif, dan psikomotorik. Oleh karena itu, instrumen atau alat yang digunakan untuk mengukur kemampuan siswa, baik dalam aspek kognitif, afektif, maupun psikomotorik, harus disesuaikan dengan kebutuhan masing-masing aspek tersebut.

Secara umum, penilaian dapat dibagi menjadi dua jenis, yaitu tes dan non-tes. Tes dapat diartikan sebagai respons seseorang terhadap suatu pertanyaan (Widoyoko, 2013). Beberapa bentuk tes yang umum digunakan antara lain pilihan ganda dan uraian bebas. Sementara itu, bentuk non-tes meliputi bagan partisipasi, daftar cek, dan skala sikap. Sebelum mencapai tahap penilaian, penyusunan instrumen merupakan langkah krusial yang tidak dapat dipisahkan dari kompetensi dasar (KD). Dalam jurnal yang ditulis oleh Siti Mariyah berjudul "Peningkatan Kompetensi Guru

dalam Menyusun Instrumen Penilaian Hasil Belajar,” ditemukan beberapa kelemahan guru dalam menyusun instrumen penilaian, yang menyebabkan instrumen tersebut tidak valid. Beberapa faktor penyebabnya antara lain: pertama, penulisan soal tidak disertai dengan petunjuk yang jelas; kedua, rumusan soal menggunakan bahasa yang sulit dipahami siswa; dan ketiga, panjang jawaban yang tidak konsisten(Siti Mariyah, 2019).

Berdasarkan kelemahan dan kekurangan guru dalam menyusun instrumen, karakteristik awal yang memiliki peran penting dalam instrumen penilaian adalah validitas. Validitas dapat diartikan sebagai ketepatan pendapat atau interpretasi yang dihasilkan dari instrumen penilaian. Dengan kata lain, suatu instrumen penilaian dianggap valid jika instrumen tersebut mampu mengukur apa yang seharusnya diukur dan mengetahui apa yang seharusnya diketahui. Oleh karena itu, penting bagi seorang guru untuk mampu menyusun instrumen penilaian yang sesuai dan valid agar kemampuan siswa dapat diukur secara tepat. Uji validitas suatu instrumen tentu harus

memperhatikan kesesuaiannya dengan kompetensi dasar serta prinsip-prinsip penilaian.

Artikel ini bertujuan untuk menganalisis validitas instrumen penilaian dalam buku Sejarah Kebudayaan Islam (SKI) kelas 12. Di dalam buku tersebut terdapat beberapa bab materi. Untuk memfokuskan analisis validitas, peneliti membatasi penelitian hanya pada satu bab, yaitu bab V yang membahas tokoh-tokoh gerakan pembaharuan Islam. Harapan dari penelitian mini ini adalah bahwa tanggung jawab guru sangatlah berat. Selain mengajar, guru juga dibebani dengan berbagai tugas administrasi, salah satunya adalah melaksanakan evaluasi pembelajaran, baik yang bersifat harian, mingguan, bulanan, maupun tahunan. Sebelum melakukan evaluasi, guru juga harus menyusun indikator dan kisi-kisi untuk memudahkan pembuatan instrumen penilaian, dengan memperhatikan prinsip-prinsip penyusunan instrumen, yaitu validitas dan reliabilitas. Semoga artikel ini dapat memberikan bantuan dalam memahami bagaimana guru seharusnya mempersiapkan semua hal tersebut dengan baik.

B. Metode Penelitian

Penelitian ini menggunakan jenis penelitian kualitatif dengan pendekatan deskriptif-analitis. Menurut Sugiyono, penelitian kualitatif adalah penelitian yang digunakan untuk meneliti kondisi objek alamiah dimana peneliti merupakan instrumen kunci (Sugiyono, 2021). Sementara itu, pendekatan deskriptif-analitis dipilih karena penelitian ini bertujuan untuk mendeskripsikan dan menganalisis validitas instrumen asesmen pada materi tokoh-tokoh gerakan pembaruan Islam dalam buku SKI kelas 12 MA Muallimin Muhammadiyah Yogyakarta. Creswell menegaskan bahwa pendekatan deskriptif-analitis tepat digunakan untuk menganalisis dokumen atau teks dengan menguraikan secara sistematis karakteristik dan kualitasnya (Creswell & Creswell, 2018).

Sumber data dalam penelitian ini terbagi menjadi dua, yaitu sumber data primer dan sumber data sekunder. Sumber data primer dalam penelitian ini adalah buku ajar Sejarah Kebudayaan Islam Kelas 12 yang digunakan di MA Muallimin Muhammadiyah Yogyakarta, terbitan

Muallimin Press tahun 2023; instrumen asesmen tertulis pada materi tokoh-tokoh gerakan pembaruan Islam yang terdapat dalam buku ajar tersebut, meliputi soal-soal latihan, tugas, dan evaluasi; serta instrumen asesmen yang dikembangkan oleh guru SKI kelas 12 MA Muallimin Muhammadiyah Yogyakarta, seperti soal latihan harian. Sedangkan sumber data sekunder dalam penelitian ini adalah dokumen kurikulum SKI kelas 12 MA, termasuk silabus, rencana pelaksanaan pembelajaran (RPP), dan kompetensi dasar yang ditetapkan oleh Kementerian Agama RI; buku-buku referensi tentang tokoh-tokoh gerakan pembaruan Islam.

C. Hasil Penelitian dan Pembahasan

1. Pengertian Validitas

Validitas instrumen dalam dunia pembelajaran merujuk pada sejauh mana sebuah alat ukur mampu menjalankan fungsinya secara tepat dalam mengukur sesuatu yang ingin diketahui (Phafiandita dkk., 2022; Sullivan, 2011). Dengan demikian, validitas menunjukkan bahwa alat

ukur tersebut benar-benar mengukur aspek atau kompetensi yang sesuai dengan tujuan evaluasi yang telah ditentukan.

Beberapa pakar pendidikan menyamakan istilah valid dengan “sahih”, sehingga validitas diartikan sebagai kesahihan suatu instrumen(Widoyoko, 2013). Di sisi lain, ada pula yang memaknai validitas sebagai ketepatan, yaitu tingkat kecocokan antara instrumen dengan objek atau indikator yang hendak diukur(Hambali dkk., 2022).

Selain itu, validitas juga dapat dipahami sebagai suatu proses interpretatif terhadap hasil data yang diperoleh melalui prosedur tertentu. Oleh sebab itu, validitas bersifat kontekstual dan tidak mutlak, bergantung pada tujuan, situasi sosial, dan pendekatan yang digunakan dalam proses penilaian(Hadaina dkk., 2021). Validitas dalam hal ini merupakan bentuk pembenaran ilmiah terhadap data melalui bukti-bukti empiris yang dapat berupa nilai, hasil pengamatan, maupun data lainnya.

Terdapat dua unsur penting di dalam validitas itu sendiri.

Pertama, validitas menunjukkan adanya derajat (Sangat baik, baik, sedang, cukup dan rendah). Kedua, validitas selalu dihubungkan dengan sesuatu yang spesifik. Sehingga penilaian mampu mengukur kompetensi hasil belajar sesuai dengan indikator yang telah dibuat oleh guru itu sendiri(Silalahi, 2020; Suryadi, 2020).

Sebuah instrumen dianggap valid apabila dapat menampilkan informasi yang benar dan representatif dari apa yang hendak dinilai. Artinya, instrumen tersebut mampu menunjukkan hubungan yang erat antara alat ukur dengan objek ukur secara tepat dan akurat (Supriyadi, 2011). Contohnya, jika ingin mengetahui sejauh mana penguasaan siswa terhadap materi Sejarah Kebudayaan Islam, maka instrumen harus mengacu pada indikator dalam mata pelajaran tersebut, bukan mengambil indikator dari pelajaran lain. Hal ini menegaskan bahwa validitas sangat bergantung pada relevansi antara alat ukur dengan tujuan evaluasi.

2. Jenis-jenis Validitas

Validitas dalam pembelajaran tidak hanya satu bentuk, melainkan terdiri dari beberapa jenis yang masing-masing memiliki fokus dan cara pengukuran yang berbeda. Pemahaman terhadap jenis-jenis validitas ini sangat penting agar guru dan peneliti pendidikan dapat menggunakan instrumen yang tepat dan sesuai dengan tujuan pembelajaran. Berikut adalah beberapa jenis validitas yang umum digunakan:

a. Validitas Isi (Content Validity)

Validitas isi merujuk pada sejauh mana suatu instrumen mampu merepresentasikan secara menyeluruh konten atau materi yang hendak diukur (Maielfi, 2021). Pertanyaan kuncinya adalah: *Apakah butir-butir soal dalam instrumen telah mencakup semua aspek penting dari materi yang diajarkan?* Untuk memastikan validitas isi, keterlibatan ahli dalam bidang studi sangat penting. Seorang ahli dapat menilai apakah instrumen tersebut telah sesuai dengan

kompetensi dan indikator pembelajaran, serta layak untuk mengukur kemampuan peserta didik secara objektif (Valentina Sahra & Aisiah, 2023).

Instrumen yang memiliki validitas isi tinggi biasanya ditandai dengan keselarasan antara tujuan pembelajaran, indikator pencapaian, dan butir soal yang dibuat (Suyasa & Divayana, 2022). Hal ini berlaku khususnya pada instrumen berbentuk tes yang digunakan untuk mengevaluasi capaian akademik siswa. Dalam hal ini, penyusunan soal harus memperhatikan redaksi yang jelas, pemilihan distraktor (jawaban pengecoh) yang logis, dan format soal yang sesuai.

Langkah awal yang umum dilakukan dalam menjamin validitas isi adalah menyusun kisi-kisi soal. Kisi-kisi ini berfungsi sebagai pedoman dalam mengembangkan soal agar seluruh materi yang diajarkan terakomodasi dengan baik. Setelah kisi-kisi disusun, tahap

berikutnya adalah menulis butir soal berdasarkan kisi-kisi tersebut. Penyusunan ini dapat dipadukan dengan analisis rasional, yakni mencocokkan setiap butir soal dengan materi yang bersumber dari buku teks atau silabus(Surur & Rakhman, 2022).

Validitas isi juga dapat dianalisis dengan membandingkan antara isi butir soal dan konstruk yang hendak diukur. Proses ini melibatkan identifikasi aspek-aspek penting dalam tema pembelajaran, penggunaan Kata Kerja Operasional (KKO), jenis soal yang digunakan (uraian, pilihan ganda, dsb.), dan kesesuaian konteks pertanyaan. Dengan demikian, validitas isi bukan hanya aspek teknis semata, tetapi juga mencerminkan kejelasan dan ketepatan arah pengukuran yang dilakukan oleh guru atau peneliti pendidikan(Ayumi dkk., 2021).

Dalam soal uraian, validitas isi mengacu pada sejauh mana soal-soal tersebut mewakili materi yang

telah diajarkan dan sesuai dengan indikator pembelajaran. Pada soal uraian, validitas isi dinilai dari:

- 1) Apakah soal mengacu pada kompetensi dasar dan indikator yang tepat?
- 2) Apakah redaksi soal jelas, tidak ambigu, dan menuntut jawaban yang sesuai dengan tujuan pembelajaran?
- 3) Apakah soal mencakup bagian penting dari materi (tidak menyimpang atau terlalu sempit)?

b. Validitas Konstruk (Construct Validity)

Validitas konstruk dalam pembelajaran merupakan bentuk validitas yang menunjukkan sejauh mana suatu instrumen mampu mengukur konstruk atau konsep teoretis yang menjadi sasaran evaluasi(Mulyanti & Rahmania, 2022). Seperti kemampuan berpikir kritis, kreativitas, atau sikap religius. Karena konstruk adalah konsep yang bersifat abstrak dan tidak bisa diamati secara langsung, maka perancang

instrumen perlu memiliki pemahaman teoretis yang kuat untuk merumuskan butir-butir soal yang sesuai. Instrumen yang dirancang tanpa landasan teori yang jelas akan sulit menangkap esensi dari konstruk yang ingin diukur, sehingga dapat menghasilkan data yang tidak akurat atau menyesatkan dalam proses penilaian pembelajaran (Mahrunnisya, 2022).

Dalam praktiknya, validitas konstruk pada instrumen pembelajaran—baik berupa soal pilihan ganda, uraian, angket, maupun lembar observasi—dapat memiliki tingkat yang bervariasi. Validitas konstruk dapat diklasifikasikan dalam lima kategori, yaitu: sangat rendah, rendah, sedang, tinggi, dan sangat tinggi. Pengukuran tingkat validitas ini biasanya dilakukan melalui analisis ahli (expert judgment) atau uji empiris, seperti uji korelasi antarbutir dan uji konstruk melalui analisis faktor. Sebagai contoh, sebuah soal uraian dikatakan

memiliki validitas konstruk tinggi jika butir soalnya benar-benar menuntut siswa menunjukkan proses berpikir analitis sesuai dengan indikator kemampuan berpikir kritis. Sebaliknya, validitas konstruk rendah terjadi apabila soal tersebut hanya menuntut hafalan atau pemahaman dangkal, padahal konstruk yang ingin diukur bersifat kompleks dan konseptual (Kaaffah dkk., 2021).

Untuk meningkatkan validitas konstruk suatu instrumen, guru atau peneliti harus melalui tahapan penyusunan yang sistematis. Hal ini mencakup penyusunan indikator berdasarkan teori yang kuat, pembuatan kisi-kisi soal yang selaras dengan tujuan pembelajaran, serta pelibatan ahli dalam proses validasi isi dan konstruk. Selain itu, uji coba instrumen dan analisis statistik juga diperlukan untuk memastikan bahwa butir soal benar-benar mencerminkan konstruk yang ditargetkan. Dengan demikian,

instrumen pembelajaran tidak hanya bersifat sahih (valid), tetapi juga mampu menghasilkan data yang bermakna dan mendukung pencapaian kompetensi siswa secara utuh.

Dalam soal uraian, validitas konstruk mengacu pada Validitas konstruk berkaitan dengan apakah soal uraian tersebut benar-benar mengukur konstruk atau kemampuan kognitif tertentu, seperti kemampuan berpikir kritis, menganalisis, menyintesis, atau mengevaluasi. Dalam hal ini, validitas konstruk pada soal uraian dinilai dari:

- 1) Apakah soal menuntut siswa menunjukkan kemampuan berpikir tingkat tinggi?
- 2) Apakah jenis soal sesuai dengan ranah kognitif yang diukur (misalnya berdasarkan Taksonomi Bloom)?
- 3) Apakah jawaban yang diharapkan mencerminkan penguasaan konsep, bukan sekadar hafalan?

3. Analisis Validitas Isi dan Konstruk Instrumen Asesmen Buku Pelajaran SKI Kelas 12 MA Muallimin Muhammadiyah Yogyakarta

Penelitian ini bertujuan untuk menganalisis validitas isi dan validitas konstruk pada instrumen asesmen berupa lima soal uraian yang terdapat dalam buku Sejarah Kebudayaan Islam (SKI) kelas 12 MA Muallimin Muhammadiyah Yogyakarta, khususnya pada Bab V: Tokoh-Tokoh Gerakan Pembaruan Islam. Instrumen yang dianalisis difokuskan pada soal-soal yang dirancang untuk mengukur pemahaman siswa terhadap pemikiran dan kontribusi tokoh-tokoh pembaru Islam klasik dan modern.

Analisis dalam penelitian ini dilakukan berdasarkan empat aspek utama yang mencerminkan pendekatan evaluatif terhadap kualitas butir soal secara menyeluruh. Pertama, kesesuaian dengan Kompetensi Dasar (KD), yaitu menilai apakah soal-soal uraian selaras dengan rumusan KD yang tercantum dalam

kurikulum mata pelajaran Sejarah Kebudayaan Islam (SKI). Aspek ini penting untuk menjamin bahwa soal yang disusun relevan dengan tujuan pembelajaran yang telah ditetapkan oleh pemerintah melalui regulasi kurikulum.

Kedua, level kognitif berdasarkan Taksonomi Bloom revisi, yang digunakan untuk mengidentifikasi tingkat berpikir yang dituntut oleh masing-masing soal (Anifarka & Rosnawati, 2023; Lestari dkk., 2022). Taksonomi ini mencakup enam level kognitif, mulai dari mengingat (C1), memahami (C2), menerapkan (C3), menganalisis (C4), mengevaluasi (C5), hingga mencipta (C6) (Listiani & Rachmawati, 2022). Dengan demikian, analisis ini dapat menunjukkan apakah instrumen asesmen mendorong kemampuan berpikir tingkat rendah (LOTS) atau justru mengembangkan kemampuan berpikir tingkat tinggi (HOTS), yang sangat dibutuhkan dalam pembelajaran abad 21.

Ketiga, kejelasan dan struktur soal, yang mencakup aspek kebahasaan, keterbacaan, dan logika penyusunan soal. Soal

yang baik seharusnya memiliki redaksi yang jelas, tidak ambigu, dan mampu dipahami dengan mudah oleh peserta didik sesuai jenjangnya (Ariawan dkk., 2022). Struktur soal yang rapi dan sistematis juga memudahkan siswa dalam menangkap inti pertanyaan dan memberikan jawaban yang sesuai (Maharani & Wahidin, 2022).

Keempat, dimensi nilai reflektif dalam pembelajaran, yang mengukur sejauh mana soal mendorong peserta didik untuk melakukan refleksi terhadap makna pembelajaran sejarah Islam dalam kehidupan nyata (Handy, 2021; Hasudungan, 2022). Dimensi ini sangat penting dalam pendidikan karakter dan moderasi beragama, karena sejarah tidak hanya dipahami sebagai pengetahuan masa lalu, tetapi juga sebagai sumber nilai yang dapat diinternalisasi dan diterapkan dalam konteks kekinian (Hastuti dkk., 2022).

Selain keempat aspek tersebut, analisis diperluas pada dua jenis validitas utama, yaitu validitas isi (content validity) dan validitas konstruk (construct

validity). Validitas isi merujuk pada sejauh mana isi soal mencakup materi dan indikator pembelajaran secara komprehensif (Kaaffah dkk., 2021). Validitas ini umumnya dinilai melalui expert judgment, dengan melihat keterwakilan butir soal terhadap ruang lingkup pengetahuan yang seharusnya diukur dalam bab tertentu (Lukman dkk., 2023).

Sementara itu, validitas konstruk menilai sejauh mana soal tersebut mampu mengukur konstruk atau kemampuan teoretik yang dimaksud dalam tujuan pembelajaran, seperti kemampuan berpikir kritis, pemahaman historis, dan kemampuan evaluatif terhadap perkembangan pemikiran Islam (Putra & El Fahmi, 2024). Validitas konstruk berperan penting untuk memastikan bahwa asesmen benar-benar merepresentasikan kemampuan yang hendak dikembangkan melalui pembelajaran SKI.

Seluruh aspek ini akan disajikan secara sistematis dalam bentuk tabel analisis, sehingga memberikan gambaran yang lebih terukur dan objektif terhadap

kualitas instrumen asesmen yang digunakan. Penyusunan analisis ini diharapkan mampu menjadi pijakan bagi perbaikan instrumen evaluasi pembelajaran, sekaligus mendukung praktik asesmen yang tidak hanya kognitif tetapi juga holistik dan kontekstual.

Adapun butir-butir soal yang dianalisis disajikan sebagai berikut.

1. Jelaskan tentang gerakan salafiyah yang dicetuskan oleh Ibnu Taimiyah!

Tabel 1. Butir soal 1

Validitas Isi	Cukup
Soal ini memang mengarah pada materi pokok dalam SKI kelas 12, tetapi belum merinci aspek penting dari gerakan Salafiyah seperti latar belakang, tujuan, atau pengaruhnya. Karena itu, cakupannya masih terbatas.	
Validitas Konstruk	Rendah
Soal hanya menuntut siswa menyebutkan atau menjelaskan secara deskriptif, tidak sampai pada tingkat analisis, sintesis, atau refleksi. Padahal “gerakan” mengandung aspek historis dan ideologis yang bisa dianalisis lebih dalam.	
Level Kognitif	C2 (<i>Understanding</i>)
Sesuai dengan jenis pertanyaan “jelaskan”, soal ini hanya menuntut siswa memahami dan mengungkapkan kembali informasi, bukan menerapkan atau menganalisisnya.	
Kejelasan Soal	Cukup
Redaksi soal jelas tetapi umum. Tidak diarahkan pada aspek spesifik seperti “tujuan” atau “pengaruh” gerakan, sehingga jawabannya bisa terlalu luas atau terlalu sempit.	

Nilai Reflektif	Rendah
Soal tidak mendorong siswa untuk berpikir kritis atau merefleksikan dampak dari gerakan tersebut terhadap pemikiran Islam saat ini.	

2. Jelaskan Pemikiran Ibnu Taimiyah tentang perlunya membuka pintu Ijtihad!

Tabel 2. Butir soal 2

Validitas Isi	Baik
Soal sesuai dengan materi penting dalam topik pembaharuan Islam dan pemikiran Ibnu Taimiyah. Fokusnya lebih spesifik dan relevan dengan tema besar ijtihad.	
Validitas Konstruk	Sedang
Soal mulai menyentuh aspek kemampuan berpikir tingkat menengah karena menuntut pemahaman dan pengaitan gagasan, tetapi belum sepenuhnya mendorong siswa untuk menilai atau mengkritisi.	
Level Kognitif	C2–C3 (Understanding–Applying)
Soal tidak hanya menuntut pemahaman tetapi juga pengaplikasian ide ijtihad dalam konteks tertentu, walaupun belum sepenuhnya pada tingkat analisis.	
Kejelasan Soal	Cukup
Redaksi soal bisa lebih tajam dengan menambahkan konteks atau batasan ruang lingkup. Namun secara umum sudah cukup untuk dipahami siswa.	
Nilai Reflektif	Sedang
Soal ini membuka peluang refleksi, misalnya dengan mengaitkan ijtihad pada konteks masa kini, meski tidak secara eksplisit diminta.	

3. Jelaskan Pemikiran Ibnu Qayyim tentang Neo-sufisme!

Tabel 3. Butir soal 3

Validitas Isi	Cukup
Soal membahas tokoh dan konsep penting, namun istilah "neo-sufisme" cukup kompleks dan membutuhkan pengantar atau pembatasan agar tidak multitafsir, terutama untuk siswa tingkat menengah.	
Validitas Konstruk	Sedang
Soal ini potensial mengarah pada kemampuan analitis, karena membahas pemikiran, tetapi karena redaksinya kurang fokus, maka tidak semua siswa bisa menggali aspek tersebut secara tepat.	
Level Kognitif	C2–C4 (Understanding–Analyzing)
Soal mencakup pemahaman dan mengarah ke analisis, namun perlu arahan agar benar-benar menyentuh analisis pemikiran Ibnu Qayyim secara utuh.	
Kejelasan Soal	Perlu difokuskan
Istilah "neo-sufisme" perlu didefinisikan atau dijelaskan konteksnya dalam pembelajaran, karena bisa membingungkan jika tidak dijelaskan sebelumnya.	
Nilai Reflektif	Sedang
Soal berpotensi mengajak siswa berpikir mendalam, tetapi karena redaksinya kurang mengarahkan pada implikasi atau relevansi, maka nilai reflektifnya tidak maksimal.	

4. Jelaskan gerakan Muwahidin yang dicetuskan oleh Muhammad bin Abdul Wahhab!

Tabel 4. Butir soal 4

Validitas Isi	Baik
Soal sesuai dengan salah satu gerakan pembaharuan Islam yang wajib dipelajari. Aspek "gerakan Muwahidin" adalah inti dari	

pemikiran Muhammad bin Abdul Wahab.

Validitas Konstruk	Sedang
Soal hanya menuntut deskripsi dan pemahaman dasar, belum sampai pada analisis terhadap pemikiran, latar belakang, atau dampak gerakan tersebut.	

Level Kognitif	C2 (<i>Understanding</i>)
Kata "jelaskan" kembali hanya menuntut pemahaman. Belum ada tuntutan berpikir lebih dalam seperti membandingkan atau mengevaluasi.	

Kejelasan Soal	Cukup
Soal cukup jelas, namun bisa lebih spesifik jika diarahkan pada aspek ideologi, tujuan, atau pengaruh gerakan Muwahidin.	

Nilai Reflektif	Sedang
Siswa bisa diminta untuk melihat relevansi gerakan ini dengan gerakan modern, tetapi karena tidak diminta secara eksplisit, maka nilai reflektifnya belum tinggi.	

5. Apa perbedaan konsep Pan-Islamisme yang dipahami Barat dengan konsep Jami'ah Islamiyah menurut Al-Afghani!

Tabel 5. Butir soal 5

Validitas Isi	Sangat Baik
Soal sangat relevan, membahas konsep besar dalam wacana politik Islam dan mencakup dua sudut pandang yang penting: Barat dan tokoh Islam.	

Validitas Konstruk	Tinggi
Soal menuntut kemampuan menganalisis dua konsep yang berbeda secara tajam, serta mengevaluasi ide pemikiran Al-Afghani. Ini menunjukkan konstruk berpikir tingkat tinggi.	

Level Kognitif	C4-C5 (<i>Analyzing-Evaluating</i>)
Siswa harus membedakan, membandingkan, dan mengevaluasi dua pandangan ideologis—ini masuk dalam	

level kognitif tinggi.

Kejelasan Soal	Jelas
Redaksi soal spesifik dan tidak multitafsir. Fokus pada perbedaan konsep dari dua perspektif membuatnya mudah dipahami dan dijawab secara kritis.	

Nilai Reflektif	Tinggi
Soal ini secara alami mendorong siswa untuk menilai dan mempertimbangkan kembali konsep persatuan Islam dalam konteks global dan ideologi modern, sehingga memberikan ruang refleksi yang dalam.	

Hasil analisis menunjukkan bahwa soal-soal asesmen dalam buku SKI kelas 12 secara umum telah mengacu pada Kompetensi Dasar (KD) yang relevan, terutama KD 3.3 dan 4.3 yang menekankan pemahaman dan penyajian pemikiran tokoh-tokoh pembaruan Islam. Hal ini menunjukkan bahwa validitas isi dari instrumen asesmen tergolong cukup hingga sangat baik, tergantung pada cakupan dan kedalaman materi yang diukur oleh masing-masing soal.

Namun demikian, dari sisi validitas konstruk, terjadi variasi yang signifikan. Soal-soal pertama hingga keempat masih berada pada level kognitif rendah hingga menengah dalam taksonomi Bloom, yakni hanya menuntut siswa untuk menjelaskan atau mendeskripsikan (C2: Understanding), dengan sedikit unsur

aplikasi (C3) dan analisis terbatas (C4). Soal nomor lima menjadi pengecualian, karena mengharuskan siswa untuk menganalisis perbedaan konseptual antara dua perspektif ideologis—yakni Pan-Islamisme Barat dan Jami'ah Islamiyah versi Jamaluddin al-Afghani. Soal ini tidak hanya menilai pemahaman, tetapi juga kemampuan berpikir kritis dan membandingkan argumen, sehingga validitas konstruknya tinggi.

Perlu diketahui bahwa validitas konstruk yang kuat harus mampu menggambarkan keterhubungan antara pertanyaan dengan konstruk teoretis yang diukur (Purnamasari, 2023). Misalnya sikap reformis, semangat ijtihad, atau kritik terhadap kolonialisme dalam konteks gerakan pembaruan Islam. Dalam konteks ini, instrumen asesmen yang hanya menilai pemahaman literal tidak cukup untuk menangkap kedalaman pemikiran siswa terhadap substansi perubahan sosial-religius yang dibawa oleh tokoh-tokoh seperti Ibnu Taimiyah, Ibnu Qayyim, dan Jamaluddin al-Afghani.

Hasil penelitian ini menunjukkan bahwa sebagian besar instrumen asesmen dalam buku SKI kelas 12 masih terfokus pada

pengukuran pemahaman literal (C2: Understanding), seperti menjelaskan definisi gerakan atau menguraikan biografi tokoh. Padahal analisis konstruk seperti "semangat ijtihad" tidak hanya sekadar pengetahuan faktual, tetapi juga kemampuan siswa dalam menganalisis konteks historis, mengevaluasi relevansi pemikiran tokoh dalam konteks kekinian, serta merefleksikan nilai-nilai pembaruan tersebut dalam kehidupan sosial-religius. Misalnya, pertanyaan tentang pemikiran Ibnu Taimiyah mengenai ijtihad seharusnya tidak hanya menuntut siswa menghafal argumennya, tetapi juga mendorong mereka untuk mengevaluasi bagaimana prinsip ijtihad tersebut dapat diterapkan dalam menyelesaikan masalah kontemporer, seperti isu pluralisme atau moderasi beragama.

Temuan penelitian ini juga mengungkap bahwa hanya soal nomor lima yang berhasil mencapai level kognitif tinggi (C4-C5), yang membuat siswa dituntut untuk menganalisis perbedaan konsep Pan-Islamisme Barat dan Jami'ah Islamiyah al-Afghani. Soal semacam ini secara implisit mengukur aspek "kritik terhadap kolonialisme" dengan

membandingkan perspektif ideologis, sehingga validitas konstruksinya dinilai tinggi. Sebaliknya, instrumen yang hanya menguji deskripsi faktual, seperti menjelaskan gerakan Salafiyah, belum cukup untuk menangkap kedalaman pemikiran siswa terhadap dinamika sosial-religius yang melatarbelakangi gerakan tersebut. Padahal, pemikiran Ibnu Taimiyah tentang Salafiyah tidak terlepas dari upaya dekonstruksi terhadap tradisi taqlid dan respon terhadap hegemoni pemikiran pada masanya. Tanpa pertanyaan yang mendorong analisis, asesmen hanya menyentuh permukaan tanpa menggali aspek teoretis yang mendasar.

Implikasinya, pengembangan instrumen asesmen perlu mengintegrasikan dimensi kognitif yang tinggi dengan nilai reflektif (Asrofi dkk., 2025). Misalnya, pertanyaan seperti, *“Bagaimana pemikiran Jamaluddin al-Afghani tentang Pan-Islamisme dapat menjadi solusi terhadap tantangan Islamophobia di era globalisasi?”* tidak hanya mengukur pemahaman historis, tetapi juga kemampuan siswa dalam menghubungkan konsep teoretis dengan realitas kontemporer.

Sesuatu yang esensial dalam membangun validitas konstruk yang holistik. Dengan demikian, instrumen asesmen tidak sekadar menjadi alat ukur pengetahuan, tetapi juga sarana untuk mengembangkan kemampuan berpikir kritis dan internalisasi nilai-nilai pembaruan Islam yang transformatif.

Selain itu, dimensi nilai (*value dimension*) dalam asesmen masih belum tergali maksimal. Hanya soal nomor lima yang mengarah pada integrasi aspek afektif dan historis-politik yang dapat mendorong siswa untuk merefleksikan peran Islam dalam membangun kesadaran global umat. Selebihnya, soal cenderung bersifat tekstual dan faktual. Padahal, dalam konteks pembelajaran karakter dan moderasi beragama, sangat penting bagi asesmen untuk menggali dimensi moral, sikap, dan nilai dari ide-ide tokoh pembaruan Islam.

Temuan ini sejalan dengan penelitian Giati Anisah, yang mengungkapkan bahwa sebagian besar instrumen yang disusun guru belum mengakomodasi prinsip validitas secara menyeluruh, baik dari aspek isi, konstruk, maupun ranah afektif (Anisah & Amreta, 2023).

Keterbatasan ini juga disebabkan oleh kurangnya pelatihan dalam penyusunan soal berbasis higher-order thinking skills (HOTS), serta terbatasnya pemahaman terhadap taksonomi kognitif dan nilai-nilai yang ingin dikembangkan dalam pendidikan Islam kontemporer.

Dengan demikian, agar instrumen asesmen benar-benar efektif sebagai alat ukur capaian pembelajaran dan pembentukan karakter, maka guru perlu:

1. Mengembangkan soal yang menggabungkan aspek kognitif, afektif, dan reflektif, khususnya dalam tema-tema tokoh pembaruan Islam.
2. Menyusun kisi-kisi soal dan rubrik penilaian yang sistematis dan mengacu pada indikator hasil belajar.

Dengan pendekatan ini, instrumen asesmen tidak hanya mengukur apa yang siswa tahu, tetapi juga bagaimana mereka berpikir, apa yang mereka nilai, dan sejauh mana mereka mampu menginternalisasi semangat pembaruan Islam dalam konteks kekinian.

E. Kesimpulan

Penelitian tentang analisis validitas instrumen asesmen pada materi tokoh-tokoh gerakan pembaruan Islam dalam buku SKI kelas 12 MA Muallimin Muhammadiyah Yogyakarta menunjukkan bahwa sebagian besar soal-soal yang dianalisis memiliki validitas isi yang cukup hingga sangat baik, namun validitas konstruknya bervariasi dari rendah hingga tinggi. Hasil analisis mengungkapkan bahwa mayoritas instrumen masih berfokus pada level kognitif rendah hingga menengah (C2-C3) yang hanya menuntut siswa untuk menjelaskan atau mendeskripsikan, dengan hanya satu soal yang mencapai level kognitif tinggi (C4-C5). Temuan ini mengindikasikan perlunya pengembangan instrumen asesmen yang tidak hanya mengukur pemahaman faktual, tetapi juga mendorong kemampuan berpikir kritis, analitis, dan reflektif siswa.

Implikasi dari penelitian ini adalah bahwa guru perlu mengembangkan instrumen asesmen yang mengintegrasikan aspek kognitif, afektif, dan nilai reflektif, khususnya dalam pembelajaran tokoh-tokoh pembaruan Islam.

Instrumen asesmen yang efektif seharusnya tidak hanya mengukur apa yang siswa ketahui, tetapi juga bagaimana mereka berpikir kritis, nilai-nilai yang mereka internalisasi, dan kemampuan mereka dalam mengaitkan pemikiran tokoh-tokoh sejarah dengan konteks kontemporer. Dengan pendekatan ini, pembelajaran Sejarah Kebudayaan Islam dapat lebih bermakna dan berkontribusi pada pembentukan karakter serta moderasi beragama peserta didik.

DAFTAR PUSTAKA

- Anifarka, A., & Rosnawati, R. (2023). Analisis Buku Teks Matematika SMP Berdasarkan Tingkat Kognitif pada Taksonomi Bloom Revisi dan Numerasi pada AKM. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 7(3), 2151–2166. <https://doi.org/10.31004/cendekia.v7i3.1701>
- Anisah, G., & Amreta, M. Y. (2023). Pengembangan instrument assessment as learning berbasis projek untuk pembelajaran menyimak dan berbicara monologis dialogis bermuatan karakter Bhineka Tunggal Ika. *KEMBARA Journal of Scientific Language Literature and Teaching*, 9(1), 76–88. <https://doi.org/10.22219/kembara.v9i1.22701>
- Ariawan, R., Zetriuslita, Z., Anggara, R. P., & Winanda, S. V. (2022). Pelatihan Penyusunan Soal Hots Bagi Guru Matematika. *Jurnal Altifani Penelitian dan Pengabdian kepada Masyarakat*, 2(1), 65–74. <https://doi.org/10.25008/altifani.v2i1.207>
- Arifin, Z. (2015). *Evaluasi Pembelajaran* Penulis. Direktorat Jenderal Pendidikan Islam Kementerian Agama RI.
- Asrofi, A., Hamilaturroyya, H., & Purwoko, P. (2025). Asesmen Pembelajaran Profetik Dalam Pendidikan Islam: Strategi Holistik Untuk Penguatan Nilai Spiritual Dan Karakter Peserta Didik. *STRATEGY: Jurnal Inovasi Strategi dan Model Pembelajaran*, 5(2), 9–21. <https://doi.org/10.51878/strategi.v5i2.4839>
- Ayumi, A. Y., Sudiyati, C. A., Daulay, S. N., Wagiran, W., & Luriawati, D. (2021). Rekonstruksi Soal Penilaian Aspek Keterampilan Reseptif Mata Pelajaran Bahasa Indonesia Kelas Vii Tahun Pelajaran 2018/2019 Smp Semesta Semarang. *Asas: Jurnal Sastra*, 2(3), 52. <https://doi.org/10.24114/ajs.v10i2.26257>
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (6th Edition). SAGE Publications.

- Hadaina, N., Widiani, I. W., & Astawan, I. G. (2021). Pengembangan Instrumen Kemampuan Kerjasama Anak Kelompok B. *Journal for Lesson and Learning Studies*, 4(1).
<https://doi.org/10.23887/jlls.v4i1.31116>
- Hambali, B., Ma'mun, A., Susetyo, B., Hidayat, Y., & Gumilar, A. (2022). Validitas dan Reliabilitas Rubrik Penilaian Tes Hasil Belajar Keterampilan High Service Untuk Siswa Sekolah Dasar. *TEGAR: Journal of Teaching Physical Education in Elementary School*, 5(2).
<https://doi.org/10.17509/tegar.v5i2.46208>
- Handy, M. R. N. (2021). Pembelajaran Sejarah Dalam Membangun Historical Awareness dan Sikap Nasionalisme Pada Peserta Didik. *Prabayaksa: Journal of History Education*, 1(1), 49.
<https://doi.org/10.20527/prb.v1i1.2196>
- Hastuti, I., Asmawulan, T., & ... (2022). Asesmen PAUD Berdasar Konsep Merdeka Belajar Merdeka Bermain di PAUD Inklusi Saymara. *Jurnal Obsesi: Jurnal ...*, Query date: 2025-02-25 14:48:54.
<https://pdfs.semanticscholar.org/f686/b616f613f5033fe6ae6d050c6da7d800b676.pdf>
- Hasudungan, A. N. (2022). Paradigma Pedagogi Reflektif (PPR): Konstruksi Nilai-Nilai Karakter dalam Materi Sejarah Maritim Untuk Sekolah Menengah Atas. *Education & Learning*, 2(2), 64–73.
<https://doi.org/10.57251/el.v2i2.347>
- Kaaffah, R. R. S., Wijiyono, A. W., & Rahmayanti, I. (2021). Validitas Isi Pada Alat Evaluasi Buku Teks Pelajaran Bahasa Indonesia Kelas X SMA. *Imajeri: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 3(2), 158–167.
<https://doi.org/10.22236/imajeri.v3i2.6572>
- Kemendikbud. (2020). *Buku Saku Asesmen Diagnosis Kognitif Berkala*. Badan Penelitian dan Pengembangan dan Perbukuan.
- Lestari, G. P., Yensy, N. A., Hanifah, H., Agustinsa, R., & Susanto, E. (2022). Analisis Soal Latihan Buku Matematika Kurikulum 2013 Kelas Viii Materi Lingkaran Berdasarkan Taksonomi Bloom. *Jurnal Penelitian Pembelajaran Matematika Sekolah (JP2MS)*, 6(1), 23–31.
<https://doi.org/10.33369/jp2ms.6.1.23-31>
- Listiani, W., & Rachmawati, R. (2022). Transformasi Taksonomi Bloom dalam Evaluasi Pembelajaran Berbasis HOTS. *JURNAL JENDELA PENDIDIKAN*, 2(03), 397–402.
<https://doi.org/10.57008/jjp.v2i03.266>

- Lukman, H., Setiani, A., & Agustiani, N. (2023). Pengembangan instrumen tes kemampuan pemecahan masalah matematis berdasarkan teori krulik dan rudnick: Analisis validitas konten. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, Query date: 2025-02-25 14:48:54. <https://doi.org/10.31004/edukatif.v3i2.316>
- Magdalena, I., Septiani, R., Ilmah, S. N., & Faridah, D. N. (2020). Analisis Kompetensi Guru dalam Proses Pelaksanaan Evaluasi Pembelajaran di SDN Peninggilan 05. *NUSANTARA*, 2, 262–275. <https://doi.org/10.36088/nusantara.v2i2.814>
- Maharani, B., & Wahidin, W. (2022). Analisis Kemampuan Literasi Peserta Didik Sekolah Dasar dalam Menyelesaikan Soal Asesmen Kompetensi Minimum. *Jurnal Basicedu*, 6(4), 5656–5663. <https://doi.org/10.31004/basicedu.v6i4.3097>
- Mahrurnisya, D. (2022). Analisis Instrumen dalam Evaluasi Pembelajaran di Sekolah. *Journal of Social Education*, 3(2), 92–98. <https://doi.org/10.23960/JIPS/v3i2.92-98>
- Maiefi, D. (2021). Validitas Instrument Penerapan Model Quantum Teaching Tipe Tandur terhadap Keterampilan Komunikasi dan Kolaborasi Mahasiswa. *EDUKATIF: JURNAL ILMU PENDIDIKAN*, 3(2), 404–411. <https://doi.org/10.31004/edukatif.v6i3.6601>
- Mulyanti, S., & Rahmania, S. (2022). Pengembangan Instrumen Tes Penguasaan Konsep Senyawa Alkil Halida: Analisis Validitas Model RASCH. *Jurnal Zarah*, 10(1), 21–27. <https://doi.org/10.31629/zarah.v10i1.4161>
- Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016 Tentang Standar Penilaian Pendidikan. (2016).
- Phafiandita, A. N., Permadani, A., Pradani, A. S., & Wahyudi, M. I. (2022). Urgensi Evaluasi Pembelajaran di Kelas. *JIRA: Jurnal Inovasi dan Riset Akademik*, 3(2), 111–121. <https://doi.org/10.47387/jira.v3i2.262>
- Purnamasari, M. I. (2023). Analisis Validitas Konstruk, Konvergen, dan Diskriminan dari Instrument Well-Being. *Jurnal Penelitian Pendidikan*, 15(1), 48–55. <https://doi.org/10.21137/jpp.2023.15.1.6>
- Putra, M. D. K., & El Fahmi, E. F. F. (2024). Tes Penilaian Situasional Kecemasan Mahasiswa Jurusan Psikologi terhadap Statistika: Uji validitas konstruk. *EDUKATIF: JURNAL ILMU PENDIDIKAN*, 6(3), 2010–2022. <https://doi.org/10.31004/edukatif.v6i3.6601>
- Ratnawulan, E., & Rusdiana, H. A. (2013). *Evaluasi*

- Pembelajaran: Dengan Pendekatan Kurikulum 2013*. Pustaka Setia.
- Silalahi, S. (2020). *Evaluasi Pembelajaran*. Kita Menulis.
- Siti Mariyah. (2019). Peningkatan Kompetensi Guru Dalam Menyusun Instrumen Penilaian Hasil Belajar Melalui Supervisi Akademik Teknik Kelompok. *Ideguru: Jurnal Karya Ilmiah Guru*, 4(1), 1–9. <https://doi.org/10.51169/ideguru.v4i1.62>
- Sugiyono. (2021). *Metode Penelitian Pendidikan*. Alfabeta.
- Sukardi. (2015). *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Bumi Aksara.
- Sullivan, G. M. (2011). A Primer on the Validity of Assessment Instruments. *Journal of Graduate Medical Education*, 3(2), 119–120. <https://doi.org/10.4300/JGME-D-11-00075.1>
- Supriyadi, G. (2011). *Pengantar & Teknik Evaluasi Pembelajaran*. Intimedia.
- Surur, M., & Rakhman, F. (2022). Workshop Penyusunan Kisi Dan Soal MA Wali Songo Situbondo. *Joong-Ki: Jurnal Pengabdian Masyarakat*, 1(1), 22–28. <https://doi.org/10.56799/joongki.v1i1.59>
- Suryadi, A. (2020). *Evaluasi Pembelajaran Jilid II*. CV. Jejak.
- Sutrisno, S., Yulia, N. M., & Fithriyah, D. N. (2022). Mengembangkan Kompetensi Guru Dalam Melaksanakan Evaluasi Pembelajaran Di Era Merdeka Belajar. *ZAHRA: Research and Thought Elementary School of Islam Journal*, 3(1), 52–60. <https://doi.org/10.37812/zahra.v3i1.409>
- Suyasa, P. W. A., & Divayana, D. G. H. (2022). Instrumen Evaluasi Model Discrepancy-CSE-UCLA Dalam Rangka Menunjang Evaluasi Efektivitas Pelaksanaan Pembelajaran Synchronous. *Jurnal Pedagogi dan Pembelajaran*, 5(2), 197–207. <https://doi.org/10.23887/jp2.v5i2.48447>
- Valentina Sahra, F., & Aisiah, A. (2023). Validitas Isi Instrumen Pengukuran Literasi Sejarah. *Jurnal Family Education*, 3(3), 328–334. <https://doi.org/10.24036/jfe.v3i3.127>
- Widoyoko, E. P. (2013). *Evaluasi Program Pembelajaran*. Pustaka Pelajar.