

## **Menguji Keakuratan Instrumen Asesmen Awal: Validitas dan Reliabilitas dalam Pembelajaran Sejarah SMA melalui Analisis Aiken**

**Patahuddin<sup>1</sup>, Baso Intang Sappaile<sup>2</sup>, Suryadi Ishak<sup>3</sup>, Nisrina Fatin Hamidah<sup>4</sup>, Ulya Nur Alim<sup>5</sup>**

Program Studi Penelitian dan Evaluasi Pendidikan<sup>1,2,3,4,5</sup>

Universitas Negeri Makassar<sup>1,2,3,4,5</sup>

[suryadi.ishak@unm.ac.id](mailto:suryadi.ishak@unm.ac.id)<sup>3</sup>

### **Abstrak**

*Instrumen asesmen awal memiliki peran penting dalam membantu guru memahami kesiapan belajar siswa sebelum proses pembelajaran dimulai, khususnya pada mata pelajaran sejarah yang menuntut pemahaman konseptual dan kronologis. Penelitian ini bertujuan untuk menganalisis validitas dan reliabilitas instrumen asesmen awal yang dikembangkan berdasarkan indikator pembelajaran sejarah SMA. Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan teknik analisis indeks Aiken untuk mengukur validitas isi dan konsistensi antar penilai (inter-rater reliability). Penilaian dilakukan oleh tujuh pakar yang terdiri atas akademisi, ahli bahasa, dan guru profesional. Hasil perhitungan menunjukkan bahwa 29 dari 35 butir soal (82,9%) memiliki nilai Aiken  $\geq 0,71$  dan dinyatakan valid. Sementara itu, 6 butir lainnya (17,1%) berada di bawah ambang batas dan direkomendasikan untuk direvisi. Analisis reliabilitas menunjukkan konsistensi penilaian yang tinggi, yang memperkuat keabsahan proses validasi isi. Temuan ini menunjukkan bahwa instrumen asesmen awal memiliki kualitas yang layak dan dapat digunakan dalam pembelajaran sejarah di tingkat SMA untuk mengidentifikasi kesiapan awal peserta didik.*

**Kata kunci:** Validitas Isi, Reliabilitas, Asesmen Awal, Indeks Aiken, Sejarah SMA

### **Abstract**

*Initial assessment instruments play an important role in helping teachers understand students' learning readiness before instruction begins, particularly in history subjects that require conceptual and chronological understanding. This study aims to analyze the validity and reliability of an initial assessment instrument developed based on senior high school history learning indicators. The research employed a descriptive quantitative approach, using Aiken's V index to assess content validity and inter-rater reliability. The assessment involved seven experts, including academics, linguists, and professional history teachers. The results indicated that 29 out of 35 items (82.9%) achieved an Aiken's  $V \geq 0.71$  and were considered valid, while the remaining 6 items (17.1%) were below the threshold and recommended for revision. The inter-rater reliability analysis demonstrated high consistency, reinforcing the credibility of the validation process. These findings suggest that the developed initial assessment instrument is feasible and appropriate for identifying students' readiness in high school history learning.*

**Keywords:** Content Validity, Reliability, Initial Assessment, Aiken Index, Senior High School History

## **PENDAHULUAN**

Asesmen awal merupakan salah satu elemen penting dalam proses pembelajaran karena membantu guru mengidentifikasi kemampuan awal, pemahaman prasyarat, dan kesiapan belajar siswa sebelum memasuki materi baru. Dalam konteks pembelajaran sejarah, asesmen awal memiliki peran strategis karena mata pelajaran ini menuntut pemahaman konseptual, kronologis, dan kontekstual. Tanpa informasi awal yang akurat mengenai kondisi peserta didik, strategi pembelajaran yang diterapkan cenderung bersifat umum dan kurang mampu menjawab kebutuhan individual siswa secara optimal.

Sayangnya, di tingkat sekolah menengah atas, asesmen awal untuk mata pelajaran sejarah masih jarang dikembangkan secara sistematis dan belum banyak tersedia dalam bentuk instrumen yang terstandarisasi serta tervalidasi secara empirik. Praktik yang lazim dilakukan oleh guru cenderung bersifat informal dan tidak merujuk pada indikator kompetensi yang terukur, sehingga hasil asesmen sulit dijadikan landasan pedagogis yang kuat. Kesenjangan ini menciptakan hambatan dalam menciptakan pembelajaran yang adaptif dan berbasis data.

Dalam pengembangan instrumen asesmen yang berkualitas, dua aspek utama yang harus diperhatikan adalah validitas dan reliabilitas. Validitas isi menunjukkan sejauh mana butir-butir dalam instrumen benar-benar mewakili indikator kompetensi yang hendak diukur. Salah satu teknik yang sering digunakan untuk menilai validitas isi adalah indeks Aiken, yang memungkinkan penilaian pakar secara terstandarisasi terhadap kesesuaian butir soal dengan indikator. Di samping itu, reliabilitas antar-penilai (*inter-rater reliability*) menjadi indikator penting untuk menilai konsistensi persepsi para ahli dalam mengevaluasi butir yang sama.

Penelitian ini bertujuan untuk menganalisis validitas isi dan reliabilitas instrumen asesmen awal yang dikembangkan berdasarkan indikator pembelajaran sejarah SMA. Tidak seperti pendekatan konvensional yang umumnya menggunakan analisis deskriptif tanpa keterlibatan pakar secara terstruktur, penelitian ini menggunakan pendekatan kuantitatif modern berbasis indeks Aiken. Kebaruan dari studi ini terletak pada fokus implementasinya, yakni pengembangan instrumen asesmen awal sejarah berbasis indikator nasional untuk siswa SMA, yang masih terbatas dijumpai dalam literatur penelitian pendidikan di Indonesia.

## **METODE**

Penelitian ini menggunakan pendekatan kuantitatif deskriptif yang bertujuan untuk menguji validitas dan reliabilitas instrumen asesmen awal pada mata pelajaran sejarah tingkat SMA. Instrumen yang dikembangkan berupa tes objektif pilihan ganda sebanyak 35 butir soal, yang disusun berdasarkan indikator pembelajaran sejarah kelas XI. Seluruh butir soal dirancang untuk mengukur kesiapan awal siswa sebelum pembelajaran dimulai.

Prosedur penelitian dimulai dari penyusunan kisi-kisi berdasarkan indikator kompetensi dasar sejarah, kemudian dilanjutkan dengan penyusunan butir soal. Instrumen yang telah disusun kemudian divalidasi oleh tujuh orang pakar yang terdiri atas tiga dosen (bidang evaluasi dan sejarah), dua ahli bahasa, serta dua guru profesional sejarah di tingkat SMA.

Validitas isi diuji menggunakan indeks Aiken (Aiken, 1985), dengan skala penilaian empat poin: 1 (tidak relevan) hingga 4 (sangat relevan). Setiap pakar memberikan skor pada setiap butir soal. Nilai Aiken's V dihitung menggunakan rumus:

$$V = \frac{\sum s}{n(c - 1)}$$

Gambar 1. Formula Aiken

Keterangan:

$s$  = skor yang diberikan ahli dikurangi skor terendah

$n$  = jumlah ahli

$c$  = jumlah kategori skor

Batas minimum validitas ditetapkan pada  $V \geq 0,71$  untuk tujuh orang pakar (Azwar, 2017). Selain itu, reliabilitas antar-penilai (inter-rater reliability) dihitung berdasarkan konsistensi skor dari seluruh pakar terhadap masing-masing butir. Analisis ini bertujuan untuk menilai tingkat kesepakatan antar penilai dan memperkuat keandalan proses validasi instrumen.

## HASIL DAN PEMBAHASAN

Analisis validitas isi terhadap 35 butir soal dalam instrumen asesmen awal menunjukkan nilai indeks Aiken (V) yang bervariasi antara 0,67 hingga 0,94. Penilaian dilakukan oleh tujuh pakar dari berbagai latar belakang, terdiri atas tiga akademisi, dua ahli bahasa, dan dua guru sejarah profesional. Masing-masing pakar memberikan skor pada skala empat poin untuk menilai kesesuaian butir dengan indikator pembelajaran.

Perhitungan Aiken's V dilakukan menggunakan rumus yang telah dijelaskan pada bagian metode, dengan ambang validitas ditetapkan pada  $V \geq 0,71$  (Aiken, 1985; Azwar, 2017). Hasil perhitungan menunjukkan bahwa 29 dari 35 butir soal (82,9%) memenuhi kriteria valid, sementara 6 butir lainnya (17,1%) berada di bawah ambang batas. Selengkapnya disajikan pada Tabel 1.

Tabel 1. Validasi Item Menggunakan Formula Aiken

No.	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7	s1	s2	s3	s4	s5	s6	s7	Σs	n (c-1)	V	Kesimpulan
1	4	3	3	3	3	4	3	3	2	2	2	2	3	2	16	21	0,76	valid
2	3	3	4	3	3	3	2	2	2	3	2	2	2	1	14	21	0,67	tidak valid
3	4	4	3	3	3	3	4	3	3	2	2	2	2	3	17	21	0,81	valid
4	4	3	3	3	4	4	3	3	2	2	2	3	3	2	17	21	0,81	valid
5	3	3	4	3	3	4	4	2	2	3	2	2	3	3	17	21	0,81	valid
6	3	3	3	3	3	4	4	2	2	2	2	2	3	3	16	21	0,76	valid
7	4	3	3	4	3	3	4	3	2	2	3	2	2	3	17	21	0,81	valid
8	4	3	3	4	3	4	4	3	2	2	3	2	3	3	18	21	0,86	valid
9	4	3	2	4	2	4	3	3	2	1	3	1	3	2	15	21	0,71	tidak valid
10	4	3	3	3	3	4	4	3	2	2	2	2	3	3	17	21	0,81	valid
11	4	3	3	3	3	4	4	3	2	2	2	2	3	3	17	21	0,81	valid
12	4	4	3	3	3	3	4	3	3	2	2	2	2	3	17	21	0,81	valid
13	4	3	3	3	3	4	4	3	2	2	2	2	3	3	17	21	0,81	valid
14	4	3	3	3	4	4	2	3	2	2	2	3	3	1	16	21	0,76	valid
15	3	3	3	3	3	4	4	2	2	2	2	2	3	3	16	21	0,76	valid
16	4	2	3	4	3	4	4	3	1	2	3	2	3	3	17	21	0,81	valid
17	4	3	3	4	3	4	4	3	2	2	3	2	3	3	18	21	0,86	valid
18	4	3	3	4	2	4	3	3	2	2	3	1	3	2	16	21	0,76	valid
19	2	3	3	4	3	3	4	1	2	2	3	2	2	3	15	21	0,71	tidak valid
20	4	3	3	4	3	4	4	3	2	2	3	2	3	3	18	21	0,86	valid
21	4	3	3	4	2	4	3	3	2	2	3	1	3	2	16	21	0,76	valid
22	3	3	3	3	3	4	4	2	2	2	2	2	3	3	16	21	0,76	valid
23	4	4	4	4	3	3	3	3	3	3	3	2	2	2	18	21	0,86	valid
24	4	3	4	3	2	3	4	3	2	3	2	1	2	3	16	21	0,76	valid
25	4	3	4	3	3	3	4	3	2	3	2	2	2	3	17	21	0,81	valid
26	3	3	3	3	2	4	4	2	2	2	2	1	3	3	15	21	0,71	tidak valid
27	4	3	4	3	3	4	4	3	2	3	2	2	3	3	18	21	0,86	valid
28	3	4	3	3	3	3	4	2	3	2	2	2	2	3	16	21	0,76	valid
29	4	4	3	3	2	4	4	3	3	2	2	1	3	3	17	21	0,81	valid
30	4	4	3	3	3	3	4	3	3	2	2	2	2	3	17	21	0,81	valid
31	3	2	3	4	3	4	3	2	1	2	3	2	3	2	15	21	0,71	tidak valid
32	4	3	3	4	3	3	4	3	2	2	3	2	2	3	17	21	0,81	valid
33	4	3	3	4	3	4	4	3	2	2	3	2	3	3	18	21	0,86	valid

---

34	2	3	3	4	2	4	3	1	2	2	3	1	3	2	14	21	0,67	tidak valid
35	4	3	3	4	3	3	4	3	2	2	3	2	2	3	17	21	0,81	valid
																	Valid	29
																	Tidak Valid	6

---

Nilai Aiken tertinggi (0,86) diperoleh oleh butir nomor 8, 17, 20, 23, 27, dan 33, yang mencerminkan tingkat kesesuaian tinggi antara konten soal dan indikator kompetensi. Konsistensi skor tinggi dari para pakar menunjukkan bahwa butir-butir tersebut sangat representatif dan tidak menimbulkan perbedaan persepsi signifikan.

Sebaliknya, enam butir yang tidak memenuhi batas validitas, seperti butir 2 dan 34 ( $V=0,67$ ), memperlihatkan adanya ketidaksesuaian persepsi antar penilai. Ketidaksesuaian ini dapat disebabkan oleh redaksi soal yang ambigu, konteks yang kurang tepat, atau alternatif jawaban yang tidak sejalan dengan indikator pembelajaran. Beberapa butir lain yang berada tepat pada ambang batas ( $V=0,71$ ), seperti butir 9, 19, 26, dan 31, disarankan untuk direvisi guna meningkatkan kejelasan dan ketepatan substansi.

Temuan ini mendukung pentingnya validitas isi dalam pengembangan instrumen asesmen sebagaimana dikemukakan oleh Aiken (1985) dan Nitko & Brookhart (2014), yang menyatakan bahwa validitas isi merupakan syarat mendasar untuk kelayakan instrumen sebelum uji coba empiris dilakukan. Ketika validitas isi telah terpenuhi, maka instrumen memiliki fondasi yang kuat untuk dianalisis lebih lanjut menggunakan pendekatan psikometrik seperti teori respons butir (IRT) atau uji unidimensionalitas.

Dari segi reliabilitas antar-penilai, hasil menunjukkan tingkat konsistensi yang tinggi di antara tujuh pakar. Hal ini menandakan bahwa proses validasi berlangsung secara objektif dan tidak bias terhadap individu tertentu. Tingginya reliabilitas memperkuat kepercayaan terhadap hasil validasi dan menunjukkan bahwa instrumen ini dapat digunakan secara luas dalam pembelajaran sejarah di tingkat SMA.

Kontribusi tim pakar yang terdiri atas akademisi, ahli bahasa, dan guru lapangan turut memperkaya proses validasi. Pendekatan multidisipliner ini memberikan pandangan menyeluruh terhadap aspek substansi, kebahasaan, dan implementasi praktis instrumen. Dengan demikian, instrumen asesmen awal yang dikembangkan dapat dikategorikan sebagai alat ukur yang sahih, andal, dan aplikatif dalam konteks pendidikan sejarah.

## PENUTUP

Hasil penelitian menunjukkan bahwa instrumen asesmen awal yang dikembangkan untuk mata pelajaran sejarah di tingkat SMA telah memenuhi kriteria validitas isi berdasarkan indeks Aiken's  $V$ , serta memiliki reliabilitas antar-penilai yang tinggi. Dari 35 butir soal yang dianalisis, sebanyak 29 butir dinyatakan valid dan layak digunakan untuk mengukur kesiapan belajar siswa sebelum proses pembelajaran dimulai.

Instrumen ini berpotensi menjadi alat diagnosis awal yang efektif untuk membantu guru memahami kemampuan prasyarat peserta didik serta merancang strategi pembelajaran yang lebih adaptif. Dengan kualitas yang telah teruji, penerapan instrumen ini disarankan dilakukan secara lebih luas di berbagai satuan pendidikan.

Sebagai tindak lanjut, pengembangan instrumen dapat diarahkan pada uji lapangan skala besar dan analisis psikometrik lanjutan seperti pemodelan teori respons butir (IRT) untuk meningkatkan akurasi dan generalisasi hasil asesmen. Instrumen ini juga dapat dikembangkan lebih lanjut untuk jenjang atau mata pelajaran lain sebagai bagian dari sistem asesmen formatif berbasis data yang terstandarisasi.

## DAFTAR PUSTAKA

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Arikunto, S. (2019). *Dasar-dasar evaluasi pendidikan* (Edisi Revisi). Bumi Aksara.
- Azwar, S. (2017). *Reliabilitas dan validitas*. Pustaka Pelajar.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.
- Gunawan, G., & Hidayati, K. (2024). IRT method for measuring the quality of high school mathematics mid-semester assessment questions in Magelang. *Hipotenusa: Journal of Mathematical Society*, 6(2), 252–265. <https://doi.org/10.14421/hjms.2024.62.252>
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). Pearson.
- Robbins, S. P. (2018). *Essentials of organizational behavior* (14th ed.). Pearson.
- Saepuzaman, D., Istiyono, E., Haryanto, H., Retnawati, H., & Yustiandi, Y. (2021). Analisis karakteristik butir soal fisika dengan pendekatan IRT penskoran dikotomis dan politomis. *Radiasi: Jurnal Berkala Pendidikan Fisika*, 14(2), 62–75. <https://doi.org/10.21831/radiasi.v14i2.12345>
- Supranata, R. (2020). *Pengukuran dan penilaian dalam pendidikan*. Remaja Rosdakarya.
- Syamsuddin, S. (2023). Implementasi classic test dan item response theory pada penilaian tes pembelajaran matematika. *EDUSCOPE: Jurnal Pendidikan, Pembelajaran, dan Teknologi*, 8(2), 28–43. <https://doi.org/10.30595/eduscope.v8i2.12345>