

## Pengembangan Instrumen Penilaian Kemampuan Berfikir Kritis dalam Pembelajaran Sejarah: *Systematic Literature Review*

Ida Nur Azizah<sup>1✉</sup>, Multazam<sup>2</sup>, Djuniadi<sup>3</sup>, Wiwi isnaeni<sup>4</sup>

<sup>1,2,3,4</sup> Penelitian dan Evaluasi Pendidikan, Sekolah Pascasarjana, Universitas Negeri Semarang, Jl. Lamongan Tengah No. 2, Bendan Ngisor, Kec. Gajahmungkur, Semarang, Jawa Tengah, 50229, Indonesia

✉ Corresponding Author: [Idanurazizah219@gmail.com](mailto:Idanurazizah219@gmail.com)

Article Info	ABSTRACT
<p><b>Timeline</b></p> <p>Accepted : 30/11/2025</p> <p>Revised : 08/01/2026</p> <p>Online : 22/01/2026</p> <p>Published : 31/01/2026</p> <p><b>Keyword:</b> Critical thinking, History learning, Assessment instrument, Systematic Literature Review, 21st-century skills</p>	<p><i>This study aims to systematically review various studies focusing on the development of critical thinking assessment instruments in history learning. The approach used is a Systematic Literature Review (SLR) by analyzing 23 national and international research articles published between 2019 and 2025. The selection process was carried out through the stages of identification, screening, eligibility, and synthesis using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) criteria. The results showed that most critical thinking assessment instruments in history learning were developed using the Research and Development (R&amp;D) model, with expert validation and reliability testing techniques involving both quantitative and qualitative approaches. The most frequently measured aspects of critical thinking include interpretation, analysis, evaluation, and inference. In addition, the latest trends show the use of digital technology and project-based authentic assessments to measure students' higher-order thinking skills. This study emphasizes the importance of developing contextual, valid, and reliable assessment instruments to support history learning oriented towards 21st-century skills.</i></p>
<p><b>Kata Kunci:</b> Berpikir kritis, Pembelajaran sejarah, Instrumen penilaian, Systematic Literature Review, Keterampilan abad ke-21</p>	<p><b>ABSTRAK</b> Penelitian ini bertujuan untuk mengkaji secara sistematis berbagai studi yang berfokus pada pengembangan instrumen penilaian kemampuan berpikir kritis dalam pembelajaran sejarah. Pendekatan yang digunakan adalah <i>Systematic Literature Review</i> (SLR) dengan menganalisis 23 artikel penelitian Nasional dan Internasional yang terbit antara tahun 2019 hingga 2025. Proses seleksi dilakukan melalui tahapan identifikasi, penyaringan, kelayakan, dan sintesis menggunakan kriteria atau pedoman pelaporan</p>



**Jurnal Pendidikan Sejarah**

*work is licensed under a*

[Creative Commons](#)

[Attribution 4.0 International](#)

[License.](#)

---

tinjauan sistematis dan meta analisis PRISMA. Hasil penelitian menunjukkan bahwa sebagian besar instrumen penilaian berpikir kritis dalam pembelajaran sejarah dikembangkan menggunakan model *Research and Development (R&D)*, dengan teknik validasi ahli dan uji reliabilitas yang melibatkan pendekatan kuantitatif maupun kualitatif. Aspek berpikir kritis yang paling sering diukur meliputi interpretasi, analisis, evaluasi, dan inferensi. Selain itu, tren terbaru menunjukkan penggunaan teknologi digital dan asesmen autentik berbasis proyek untuk mengukur kemampuan berpikir tingkat tinggi siswa. Penelitian ini menegaskan pentingnya pengembangan instrumen penilaian yang kontekstual, valid, dan reliabel agar mampu mendukung pembelajaran sejarah yang berorientasi pada keterampilan abad ke-21.

## PENDAHULUAN

Kurikulum merupakan seperangkat pedoman penyelenggaraan pendidikan. Kurikulum dipandang sebagai suatu rencana yang disusun untuk melancarkan proses belajar mengajar di bawah bimbingan dan tanggung jawab sekolah atau lembaga pendidikan beserta staf pengajarnya (Fauzan, 2017: 58). Kurikulum diharapkan mampu menjadi acuan dalam pengembangan potensi siswa, baik pengetahuan, keterampilan maupun karakter. Kurikulum bersifat dinamis dan selalu mengalami pembaruan sesuai dengan perkembangan zaman. Indonesia saat ini menerapkan kurikulum merdeka sebagai kurikulum Nasional. Kurikulum ini menekankan pada penguatan karakter dan keterampilan abad 21. Pesatnya perkembangan dibidang teknologi, informasi, ekonomi, sosial, dan budaya mengarahkan pendidikan untuk tidak hanya menekankan pada aspek pengetahuan saja. Namun karakter dan kompetensi yang relevan sangat dibutuhkan agar mampu beradaptasi dalam dunia yang lebih kompleks.

Keterampilan abad 21 merupakan kompetensi yang terdiri dari berfikir kritis, kreatif, kolaborasi, dan komunikasi. Melalui integrasi kompetensi tersebut diharapkan tercipta Indonesia yang berakhlaq mulia, cerdas, kreatif, dan berdaya saing global. Berfikir kritis sangat diperlukan dalam menghadapi arus informasi dan meningkatkan kemampuan pemecahan masalah melalui analisis, menilai, dan mengevaluasi sehingga menghasilkan sebuah keputusan yang berkualitas. Dalam pembelajaran sejarah, berfikir kritis tidak hanya sekedar memahami peristiwa yang telah terjadi, tetapi juga menganalisis sebuah fakta, menafsirkan sumber (*heuristik*), menganalisis keaslian sumber (*verifikasi*) dan kredibilitas, menafsirkan makna (*interpretasi*), serta membangun argumentasi.

Berdasarkan hasil PISA (2022), keterampilan berfikir tingkat tinggi Indonesia masih rendah. 99 % murid Indonesia hanya mampu menjawab soal dengan tingkah sedang (*Lower Order Thinking Skills*). Berbagai kajian menunjukkan bahwa kemampuan berfikir kritis murid di Indonesia masih relatif rendah, terutama mata pelajaran sejarah dengan dominasi hafalan (Kemdikbud, 2023; Facione, 2015). Untuk itu, dilakukanlah berbagai pengembangan instrumen untuk meningkatkan kompetensi tersebut. Misalnya, R. Herpiana, U. Rosidin, Abdurrahman (2019) yang mengembangkan instrumen penilaian untuk melatih berpikir kritis dan kreatif dalam pembelajaran fisika SMA. Penelitian ini menghasilkan instrumen layak (validasi ahli 72%), praktis (83.3%), dan efektif meningkatkan kemampuan berpikir kritis & kreatif siswa (56–57%).

Abidin & Fatimatuzzaro (2025) juga mengembangkan instrumen asesmen keterampilan berpikir kritis pada materi difusi dan osmosis (mata pelajaran biologi). Hasil penelitiannya ialah 20 dari 24 butir soal valid;

reliabilitas sedang (0.672); serta instrumen membantu guru menilai kemampuan berpikir kritis siswa.

Berdasarkan elaborasi diatas, penelitian ini dilakukan dengan tujuan untuk mengetahui:

- 1) Jenis instrumen apa saja yang telah dikembangkan untuk mengukur kemampuan berfikir kritis dalam pembelajaran sejarah?
- 2) Kerangka konseptual apa yang sering digunakan untuk pengembangan dan validasi instrumen berfikir kritis?
- 3) Bagaimana validitas dan reliabilitas instrumen yang dikembangkan?
- 4) Variabel apa saja yang telah diidentifikasi sebagai faktor yang terukur terkait kemampuan berfikir kritis dalam pembelajaran sejarah?

## **METODE**

Penelitian ini menggunakan metode penelitian *Systematic Literature Review* (SLR). Sebuah penelitian sistematis yang bertujuan untuk mengumpulkan, mengevaluasi secara kritis, mengintegrasikan, serta menyajikan temuan dari berbagai studi penelitian terkait topik.

### **Jenis metode atau jenis pendekatan**

Pada metode SLR, review dilakukan berdasarkan RQ yang terformulasikan dengan jelas, sistematis dan eksplisit dalam mengidentifikasi, menyeleksi, dan menilai secara kritis publikasi primer, mengekstrak dan mensintesis data untuk menjawab RQ. Berikut tahapan SLR yang dilakukan untuk pengumpulan data dan prosedur analisis data:

- 1) *Planning (Develop Protokol Adn Research Question)*

Langkah awal yang dilakukan pada tahap ini ialah menentukan tujuan dan protokol SLR serta merumuskan *Research Questions (RQ)* yang jelas. Kerangka yang digunakan ialah kerangka *Population* (Siswa SMK),

*Intervention* (pengembangan instrumen berfikir kritis), *Comparison* (Asesmen berbasis kertas dan digital), *Outcomes* (validitas dan reliabilitas), *Study Characteristics* (studi pengembangan, artikel empiris).

2) *Collect Paper* (Penelusuran Literature)

Dilakukan dengan identifikasi kata kunci, terms, dan sintak serta melakukan penelusuran di database Google.scholar. Pada tahap ini, penulis menggunakan Keyword yang sesuai dengan pengembangan yang akan dilakukan dengan menggunakan *Publish or Perish* (PoP).

3) *Select Egilible Study* (Seleksi Studi yang layak)

Pada tahap ini, dilakukan screening judul, screening abstrak, serta screening full-text menggunakan kriteria inklusi dan eksklusi dalam proses seleksi. Penulis menggunakan aplikasi Zetero.

4) *Acces Study Quality* (Menilai Kualitas Studi)

Pada tahap ini, dilakukan evaluasi kualitas metodologi artikel terpilih dengan menggunakan instrumen pertanyaan yang berpedoman pada Joana Briggs Institute. Tujuannya ialah agar validitas dalam penilaian berkualitas tinggi. Jenis penelitian pengembangan instrumen berfikir kritis yang dilakukan ialah pengembangan (R&D), sehingga alat yang relevan ialah *Quasi-Eksperimental Studies*. Analisis dilakukan secara manual pada excel dengan kategori kualitas rendah (0-40%0, sedang (41-70%), dan tinggi (71-100%).

5) *Extrac data*

Pembuatan form ekstraksi data sesuai RQ meliputi penulis, tahun, tujuan, metode penelitian, jumlah sampel, temuan utama, dan keterbatasan. Pada tahap ini, dilakukan secara manual pada aplikasi excel sehingga didapatkan data sesuai Research Question.

#### 6) Analysis & Syntesis

Mengategorikan data hasil ekstrak baik kualitatif atau kuantitatif kemudian menjawab RQ dengan menemukan pola, tren, gap penelitian, dan rekomendasi. Pada tahap ini, data yang diperoleh dari studi terpilih dianalisis.

#### 7) Manuscript Presentation

Menyusun laporan akhir dengan mengikuti pedoman pelaporan secara transparan yaitu *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA).

#### **Prosedur pengumpulan data**

Tujuan dari pengumpulan data ialah untuk mengidentifikasi dan mengumpulkan semua studi yang relevan dengan penelitian yang akan dilakukan (Thomas & Harden, 2008). Pengumpulan data dilakukan dengan menggunakan perpustakaan digital, antara lain; mengumpulkan jurnal yang relevan dari database *google.scholar* melalui aplikasi *Publish or Perish*, perumusan kata kunci (*keyword*), melakukan seleksi inklusi dan eksklusi menggunakan aplikasi *zetero* dan manual.

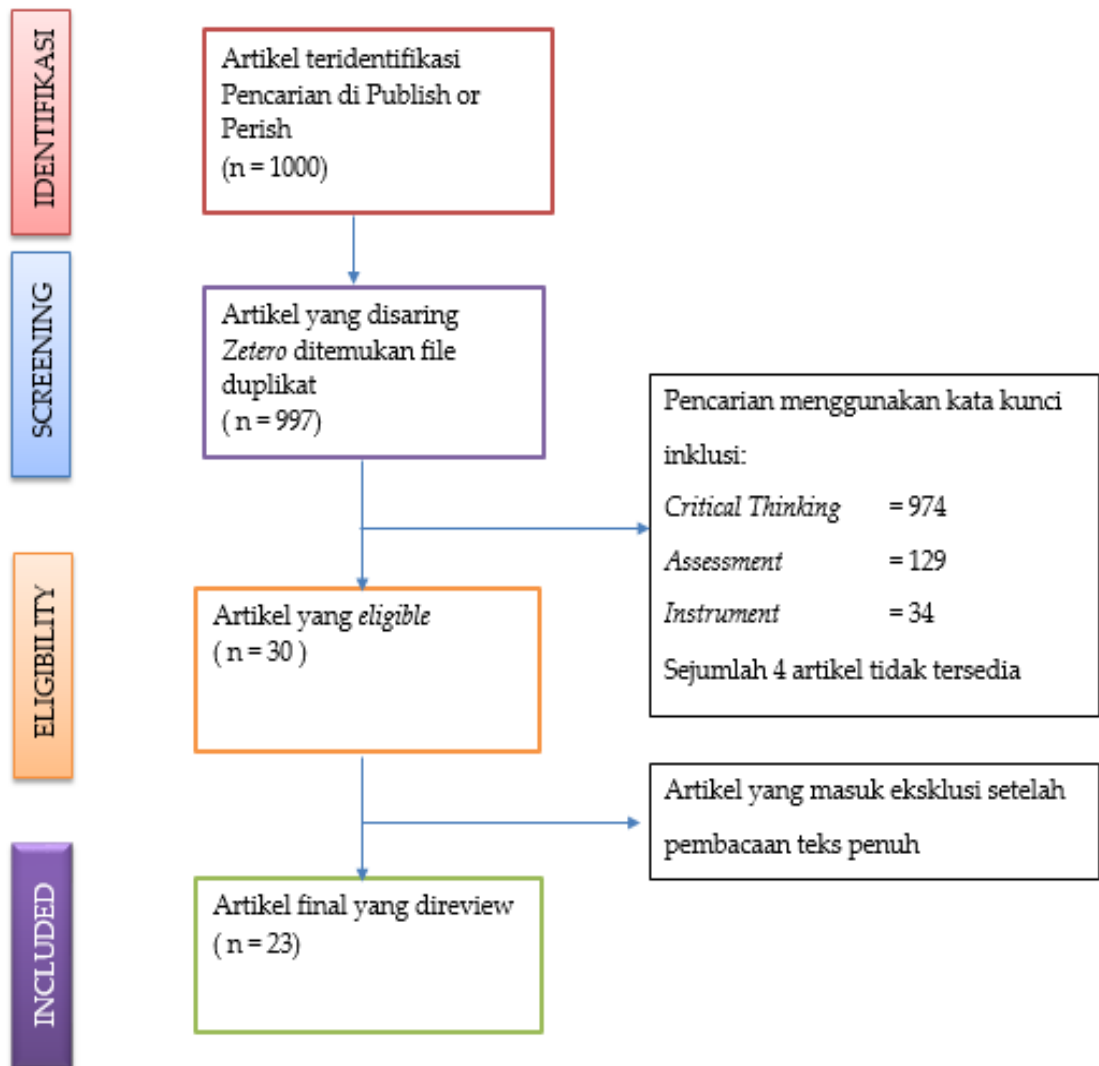
#### **Prosedur analisis data.**

Prosedur analisis data dilakukan secara sistematis berdasarkan langkah pada SLR. Berdasarkan keywords dan pembatasan tahun 2019-2025 dan maksimum result 1000, didapatkan 997 paper (*Publish or Perish*). Selanjutnya masuk pada screening judul dan abstrak di aplikasi *zetero*, dengan keyword "critical thinking" dihasilkan 974 paper yang sesuai, ditambahkan dengan "Assesment" (dihasilkan 129 paper yang sesuai), ditambahkan dengan "Instrument" (dihasilkan 34 paper yang sesuai). Kemudian screening full-text sehingga dilakukan cara untuk mendapatkan

teks secara utuh dengan langkah blok semua paper terpilih (34), klik kanan, pilih find full text. Maka akan didapatkan keterangan dari semua paper tersebut. Seperti abstrak, penulis, jenis item, judul, publikasi, tahun, URL, serta file yang sudah terdownload oleh zetero. Kemudian dilakukan kembali filter yang menghasilkan 30 paper. Screening terakhir menghasilkan paper sebanyak 23 karena ditemukan batasan yang kurang sesuai serta pengembangan yang kurang sesuai.

## **HASIL**

Penentuan artikel final yang akan dianalisis berasal dari proses identifikasi terhadap 1000 artikel. Hasil proses ini diidentifikasi menggunakan Publish or Perish. Selanjutnya dilakukan screening artikel menggunakan *zetero* dan ditemukan 3 file duplikat. Pada proses penyaringan artikel yang relevan secara keseluruhan diperoleh data sejumlah 34 artikel. Namun setelah dilakukan pencarian di *zetero* dan manual terdapat 4 artikel tidak tersedia filenya sehingga artikel tersaring yang eligible berjumlah 30. Sementara itu, untuk menentukan hasil akhir artikel, dilakukan kembali pembacaan teks penuh untuk mengidentifikasi kesesuaian metode, jenis penelitian, tahun penelitian, variabel penelitian, dan kesesuaian pertanyaan penelitian. Seleksi artikel atau sumber data yang digunakan sebagai pendekatan dari penelitian tergambar pada diagram prisma berikut.



Gambar 1. Diagram PRISMA

## PEMBAHASAN

Proses penyaringan artikel yang dilakukan secara sistematis mengacu pada pedoman PRISMA menghasilkan 23 paper. Artikel yang memenuhi seluruh kriteria inklusi ditetapkan sebagai artikel yang disertakan (*include studies*) dalam kajian sistematis. Berikut artikel terkait yang dianalisis secara mendalam untuk dilakukan sintesis tematik dan pemetaan temuan penelitian.

**Tabel 1.** Artikel Review

No	Author(s)	Tahun	Negara	Judul	Metode	Hasil/Temuan
1	Abidin & Fatimatuz-zaro	2025	Indonesia	<i>Development of a Critical Thinking Skills Assessment Instrument for Students on Diffusion and Osmosis Subject</i>	R&D (ADDIE Model)	20 dari 24 butir soal valid; reliabilitas sedang (0.672); instrumen membantu guru menilai kemampuan berpikir kritis siswa.
2	Asiah & Festiyed	2024	Indonesia	<i>Development of Diagnostic, Formative and Summative Assessment Instruments in the PjBL Model to Stimulate Students' Critical and Creative Thinking Skills</i>	R&D (ADDIE)	Instrumen valid dan praktis; efektif meningkatkan hasil belajar dan berpikir kritis siswa (N-Gain tinggi).
3	Fairuz, T.	2021	Indonesia	<i>Assessment Instrument of Critical Thinking Skills for Student on Light Interaction with Organisms</i>	R&D (Validasi Ahli & Empiris, KR-20)	25 item soal; 18 valid tinggi-sedang; reliabilitas tinggi (KR-20); mengukur lima aspek berpikir kritis (Ennis).
4	Agustini, Rery, & Anwar	2020	Indonesia	<i>Creative Problem Solving (CPS)-Based Assessment Instrument for Critical Thinking Ability on Stoichiometry Materials</i>	R&D (Borg & Gall, 3 Tahap)	Menghasilkan 10 butir soal valid dan reliabel (r=0,749); efektif mengukur keterampilan berpikir kritis siswa.

5	Agustini et al.	2021	Indonesia	<i>The Development and Validity of CPS-Based Assessment Instrument for Critical Thinking Ability on Stoichiometry Materials</i>	R&D (Borg & Gall, 3 tahap)	Instrumen memenuhi syarat validitas isi, bahasa, dan konstruk; mendukung asesmen berpikir kritis dalam kimia.
6	M. Ikhwanul Hakim, Kartono, Wahyu Lestari	2019	Indonesia	<i>Developing Assessment Instruments of Students' Critical Thinking Ability in Trigonometry Material Mathematics Subjects</i>	Borg&Gall 9 langkah. Validasi isi (Aiken's V), reliabilitas (Ebel), CFA dengan AMOS.	Instrumen dinyatakan sangat valid (Aiken's $V > 0.3$ ), reliabel ( $\alpha=0.783$ ), praktis (skor 85.6), sehingga layak digunakan guru matematika.
7	Mir'atul Hasanatin, Eli Rohaeti	2021	Indonesia	<i>How are the Characteristics of Integrated Assessment Instruments to Measure Critical Thinking Skills and Scientific Attitudes of High School Students?</i>	4D data dianalisis dengan PCM 1-PL (Rasch).	Nilai validasi Aiken's $V = 0.93$ ; reliabilitas item = 0.96; person = 0.71; semua item fit model PCM 1-PL sehingga instrumen layak digunakan.
8	R. Herpiana, U. Rosidin, Abdurrahman	2019	Indonesia	<i>Development of Instruments to Train Critical and Creative Thinking Skills in Physics Assessment for High School Students' Learning</i>	R&D (Borg & Gall 7 langkah), validasi ahli, uji terbatas & luas.	Instrumen layak (validasi ahli 72%), praktis (83.3%), dan efektif meningkatkan kemampuan berpikir kritis & kreatif siswa (56–57%).
9	Junil Adri, Refdinal, Ambiyar, Arman Shah Abdullah	2022	Indonesia–Malaysia	<i>Critical Thinking Skills in Performance-Based Assessment: Instrument Development and Validation</i>	R&D Borg & Gall (10 langkah), validasi konten, bahasa, dan konstruk.	Instrumen dinyatakan valid dan layak digunakan, mampu menstimulasi kemampuan berpikir kritis dalam praktik pembelajaran vokasi.
10	Ryzka Cahyaningrum, Wahyu Lestari, Supriyadi	2022	Indonesia	<i>Needs Analysis of Development of Critical Thinking Skills Assessment Instruments on Vocational School English Subjects</i>	Kualitatif, studi kasus, dengan observasi, wawancara, dokumentasi (triangulasi).	Sekolah memiliki fasilitas teknologi, tetapi guru belum memiliki instrumen berpikir kritis sehingga dibutuhkan pengembangan instrumen berbasis teknologi untuk asesmen.

11	Muhammad Minan Chusni, Suherman Suherman	2021	Indonesia-Hungaria	<i>Developing Authentic Assessment Instrument Based on Multiple Representations to Measure Students' Critical Thinking Skills</i>	4-D, Analisis Rasch, CVI, IRT.	Aiken's V = 1.00 (sangat valid), reliabilitas item 0.97, person 0.86. Instrumen sesuai PCM 1-PL dan efektif mengukur keterampilan berpikir kritis siswa.
12	Erina Septenia Holiqi, Maryani, Sri Handono Budi Prastowo	2021	Indonesia	<i>Development of Contextual Assessment Instruments for Students' Critical Thinking Skills Using Kahoot Quiz</i>	R&D, uji terbatas (10)&luas (30), uji validasi isi, reliabilitas, tingkat kesukaran, daya beda.	Instrumen sangat layak (validasi ahli materi 93.33%, media 81.05%), respon siswa positif, nilai rata-rata berpikir kritis 64 (kategori baik).
13	Dian Kurniati, Dinawati Trapsilasiwi, Abdur Rahman As'ari, Hasan Basri, Sharifah Osman	2022	Indonesia & Malaysia	<i>Prospective Mathematics Teachers' Critical Thinking Disposition in Designing Cognitive and Psychomotor Assessment Instruments</i>	Kualitatif deskriptif (eksploratori)	Mahasiswa menunjukkan disposisi berpikir kritis berupa keingintahuan, kedewasaan, keterbukaan, dan pencarian kebenaran selama pembelajaran kolaboratif.
14	Tri Lestari, Undang Rosidin, I Wayan Distrik, Chandra Ertikanto, Kartini Herlina	2024	Indonesia	<i>Analysis of Assessment Instrument Needs to Measure Students' Critical Thinking Skills and Self-Efficacy in Physics Learning</i>	Survei (Google Form)	45% guru tidak pernah menilai berpikir kritis; 60% tidak menilai efikasi diri; kebutuhan instrumen sangat tinggi.
15	Endang M. Kurnianti, Iva Sarifah	2023	Indonesia	<i>The Development of Assessment Instruments of Critical Thinking Ability in Mathematics Learning in Elementary School</i>	Research and Development (Borg & Gall, Dick & Carey)	Instrumen dinyatakan valid, reliabel, dan layak digunakan.

16	Rusmini, Suyono, Rudiana Agustini	2021	Indonesia	<i>Development of Critical Thinking Skills and Argumentation Skills Assessment Instruments Towards 21st-Century Skills Based on Non-Routine Problems</i>	Research & Development	Instrumen sangat valid, reliabilitas rendah; siswa sadar keterbatasan berpikir kritis mereka.
17	Y. Yokhebed, Remy M.D. Karmadi, Luvia R. Nastiti	2025	Indonesia	<i>Validity and Reliability Analysis of a Socioscientific Issues-Based Critical Thinking Self-Assessment Instrument Using the Rasch Model</i>	Kuantitatif (Rasch model)	Cronbach's alpha tinggi, validitas item baik, reliabilitas instrumen sangat baik.
18	Rohmatul Maghfiroh, Sri Wahyuni, Zainur Rasyid Ridlo	2023	Indonesia	<i>The Development of Critical Thinking Skills Assessment Instrument Based on Nearpod in Junior High School Science Learning</i>	R&D (GLAI Model)	Valid (81.7%), reliabel (0.602), praktis (97.66%), efektif mengukur berpikir kritis.
19	W. Isnaeni, E. Rudyatmi, S. Ridlo, S. Ingesti, L.R. Adiani	2021	Indonesia	<i>Improving Students' Communication Skills and Critical Thinking Ability with ICT-Oriented Problem-Based Learning and Assessment Instruments with HOTS Criteria</i>	Pre-experimental (one group pretest-posttest)	N-gain berpikir kritis 0.31 (sedang), komunikasi 83 (sangat baik).
20	Chusni & Suherman	2021	Indonesia & Hungary	<i>Developing authentic assessment instrument based on multiple representations to measure students' critical thinking skills</i>	Research and Development (4-D)	The developed essay test instrument was valid (Aiken V=1.00), reliable (0.97 items, 0.86 persons), and suitable for measuring students' critical thinking skills.
21	Amini & Arliani	2025	Indonesia	<i>Assessment Quality in Mathematics Learning: An Analysis of Critical Thinking Instruments for Derivative Topics</i>	Quantitative descriptive analysis	The items had high validity (V-Aiken=0.8889), reliability (0.924 for choice, 0.777 for essay), with minor revisions

22	Amrina et al.	2024	Indonesia	<i>Developing HOTS-Based Mathematics Assessment Instruments to Improve Critical Thinking Skills of Prospective Elementary School Teachers</i>	Research and Development (Borg & Gall Model)	required for low discriminating items. Instrument was very valid and reliable; students' HOTS average 74.65 (moderate), critical thinking 62.89% (moderate).
23	Cahyaningrum, Lestari & Supriyadi	2022	Indonesia	<i>Needs analysis of development of critical thinking skills assessment instruments on vocational school English subjects</i>	Qualitative case study (observations, interviews, documentation)	Teachers and schools lack critical thinking assessment instruments; need for technology-based tools. Students face vocabulary and grammar difficulties.

Tahapan *Systematic Literature Review* yang telah dilakukan memberikan jawaban atas pertanyaan penelitian. *Pertama*, hampir semua penelitian menggunakan model pengembangan *Research and Development* melalui uji validitas dan empiris. Jenis instrumen-instrumen yang dikembangkan untuk mengukur kemampuan berfikir kritis pada 23 artikel yang telah dianalisis sangat beragam. Antara lain:

- 1) Tes/ soal pilihan ganda beralasan untuk menilai keterampilan analisis, interpretasi, dan evaluasi.
- 2) Instrumen asesmen formatif dan sumatif untuk menilai kemampuan berfikir kritis
- 3) Instrumen berupa rubrik untuk mengukur dan menafsirkan sumber sejarah, memberikan pendapat, serta menarik sebuah kesimpulan.

*Kedua*, kerangka konseptual yang digunakan ialah pendekatan validasi teori tes klasik dengan model pengembangan ADDIE (*Analysis, Design, Development, Implementation, Evaluation*) dengan kerangka berfikir

kritis Ennis dan Facione. Kerangka ini digunakan untuk mendesain indikator dalam tes kemampuan berfikir kritis.

*Ketiga*, validitas dan reliabilitas instrumen yang dikembangkan dilakukan oleh pakar pendidikan sejarah dan pengukuran (ahli) serta siswa (empiris). Sebagian besar butir soal dalam pengembangan instrumen dinyatakan valid dengan kategori sedang sampai tinggi. Pada penelitian yang dilakukan Abidin dan Fatimatuzzahro (2025) menghasilkan 20 dari 24 butir soal valid. Sedangkan reliabilitas umumnya menunjukkan kategori sedang hingga tinggi dengan koefisien reliabilitas 20 sebesar 0,79 (Fairuz:2021).

*Keempat*, variabel yang telah diidentifikasi sebagai faktor terukur yang terkait dengan kemampuan berfikir kritis antara lain:

- 1) Sikap berfikir kritis
- 2) Kognitif berupa analisis, evaluasi, interpretasi, inferensi, eksplanasi, pengambilan keputusan
- 3) Keterampilan berfikir sejarah dengan menafsirkan sumber, menghubungkan sebab-akibat, dan membangun pernyataan berdasarkan fakta dan data sejarah.
- 4) Keterampilan berfikir tingkat tinggi

Berdasarkan hasil tersebut ditemukan pola penelitian, tren penelitian, dan gap penelitian sebagai berikut:

- 1) Pola Penelitian

Berdasarkan 23 paper yang dianalisis, pengembangan instrumen digital belum ditemukan. Sebagian besar jenis instrumen berupa tes objektif (pilihan ganda) atau uraian pendek. Hanya sebagian kecil yang menggunakan observasi, *performance assesment*, dan rubrik penilaian.

Sementara itu, sebagian besar studi dilakukan pada sekolah formal tingkat menengah dan tidak ditemukan studi yang dilakukan pada jenjang non-formal.

Kerangka teoritis yang digunakan ialah Ennis (2 paper) dan Facione (1 paper), lainnya hanya menyebutkan kerangka konseptual eksplisit. Sedangkan model pengembangan yang terdapat pada paper sebagian besar R&D dengan kerangka ADDIE yang berfokus pada analysis, design, developmen, dan uji coba terbatas. Uji validitas pada hampir seluruh paper dilakukan dengan validitas isi melalui expert judgment. Sedangkan uji reliabilitas digunakan menggunakan cronbach's alpha. Jarang dilakukan uji validitas konstruk (EFA/CFA), kriteria eksternal, dan reliabilitas temporal.

## 2) Tren

Kerangka berfikir kritis oleh facione dan ennis mulai digunakan meskipun masih terbatas. Terdapat integrasi pengembangan instrumen dengan model *Project Based Learning* (PjBL) dan *Creative Problem Solving*. Namun uji implementasi instrumen masih terbatas pada satu sekolah. Sehingga pengujian implementasi luas masih terbatas.

## 3) Gap penelitian

Analisis psikometrik lanjutan masih jarang dilakukan, seperti halnya EFA/CFA dan item analysis untuk menunjukkan performa butir. Hampir tidak ada studi yang melakukan test-retest/ konsistensi antar penilai sehingga membatasi bukti reliabilitas jangka panjang (reliabilitas temporal dan inter-rater). Sementara itu, ditemukan pula rendahnya eksplorasi pada penilaian autentik, meskipun konteksnya pembelajaran sejarah namun dimensi analisis sumber, argumentasi historis, atau

berfikir historis belum muncul. Kearifan lokal dapat dijadikan sebagai salah satu variabel dalam pembelajaran sejarah untuk memperkaya jati diri bangsa.

## KESIMPULAN

Penelitian tentang pengembangan instrumen penilaian berfikir kritis telah dilakukan oleh sebagian besar akademisi diberbagai jenjang, bahkan di negara Hungaria dan Amerika. Sebagian besar instrumen yang dikembangkan berupa soal uraian, tes pilihan ganda beralasan, serta rubrik. Instrumen yang dikembangkan menunjukkan tingkat validasi dan reliabilitas sedang hingga tinggi. Model pengembangan yang digunakan ialah model ADDIE (*Analysis, Design, Development, Implementation, Evaluation*) dengan kerangka berfikir kritis Ennis dan Facione.

Berdasarkan penelitian SLR yang dilakukan ditemukan masih adanya celah yang dapat dilakukan antara lain pemanfaatan digital, pembelajaran sejarah, variabel pengukur yang berbeda seperti model pembelajaran berbasis masalah atau pembelajaran berbasis kearifan lokal.

## DAFTAR PUSTAKA

- Abidin, A., & Fatimatuzzahro, N. (2025). Development of a critical thinking skills assessment instrument for students on diffusion and osmosis subject. *Jurnal Penelitian Pendidikan IPA*, 11(1), 45–55.
- Adri, J., Refdinal, Ambiyar, & Abdullah, A. S. (2022). Critical thinking skills in performance-based assessment: Instrument development and validation. *Journal of Physics: Conference Series*, 2309(1), 012021.
- Agustini, R., Rery, R. U., & Anwar, L. (2020). Creative problem solving (CPS)-based assessment instrument for critical thinking ability on stoichiometry materials. *Journal of Physics: Conference Series*, 1467(1), 012045.
- Agustini, R., Rery, R. U., & Anwar, L. (2021). The development and validity of CPS-based assessment instrument for critical thinking ability on

- stoichiometry materials. *Journal of Physics: Conference Series*, 1760(1), 012032.
- Amini, A., & Arliani, A. (2025). Assessment quality in mathematics learning: An analysis of critical thinking instruments for derivative topics. *Jurnal Pendidikan Matematika*, 19(1), 33–44.
- Amrina, A., Yusri, Y., & Rahman, R. (2024). Developing HOTS-based mathematics assessment instruments to improve critical thinking skills of prospective elementary school teachers. *Jurnal Pendidikan Dasar*, 15(2), 112–123.
- Asiah, A., & Festiyed, F. (2024). Development of diagnostic, formative, and summative assessment instruments in the PjBL model to stimulate students' critical and creative thinking skills. *Jurnal Pendidikan IPA Indonesia*, 13(1), 85–96.
- Cahyaningrum, R., Lestari, W., & Supriyadi, S. (2022). Needs analysis of development of critical thinking skills assessment instruments on vocational school English subjects. *Jurnal Pendidikan Vokasi*, 12(2), 157–166.
- Chusni, M. M., & Suherman, S. (2021). Developing authentic assessment instrument based on multiple representations to measure students' critical thinking skills. *Journal of Physics: Conference Series*, 1796(1), 012054.
- Fairuz, T. (2021). Assessment instrument of critical thinking skills for students on light interaction with organisms. *Journal of Physics: Conference Series*, 1842(1), 012041.
- Hakim, M. I., Kartono, & Lestari, W. (2019). Developing assessment instruments of students' critical thinking ability in trigonometry material mathematics subjects. *Journal of Research and Advances in Mathematics Education*, 4(2), 91–102.
- Hasanatin, M., & Rohaeti, E. (2021). Characteristics of integrated assessment instruments to measure critical thinking skills and scientific attitudes of high school students. *Jurnal Pendidikan Fisika Indonesia*, 17(2), 134–143.
- Herpiana, R., Rosidin, U., & Abdurrahman. (2019). Development of instruments to train critical and creative thinking skills in physics assessment for high school students' learning. *Journal of Physics: Conference Series*, 1155(1), 012032.
- Holiqi, E. S., Maryani, & Prastowo, S. H. B. (2021). Development of contextual assessment instruments for students' critical thinking

- skills using Kahoot quiz. *Journal of Physics: Conference Series*, 1832(1), 012043.
- Isnaeni, W., Rudyatmi, E., Ridlo, S., Ingesti, S., & Adiani, L. R. (2021). Improving students' communication skills and critical thinking ability with ICT-oriented problem-based learning and HOTS-based assessment instruments. *Jurnal Pendidikan IPA Indonesia*, 10(3), 412–423.
- Kurnianti, E. M., & Sarifah, I. (2023). The development of assessment instruments of critical thinking ability in mathematics learning in elementary school. *Jurnal Pendidikan Dasar*, 14(1), 55–65.
- Kurniati, D., Trapsilasiwi, D., As'ari, A. R., Basri, H., & Osman, S. (2022). Prospective mathematics teachers' critical thinking disposition in designing cognitive and psychomotor assessment instruments. *International Journal of Instruction*, 15(1), 537–552.
- Lestari, T., Rosidin, U., Distrik, I. W., Ertikanto, C., & Herlina, K. (2024). Analysis of assessment instrument needs to measure students' critical thinking skills and self-efficacy in physics learning. *Jurnal Pendidikan Fisika*, 12(2), 101–112.
- Maghfiroh, R., Wahyuni, S., & Ridlo, Z. R. (2023). Development of critical thinking skills assessment instrument based on Nearpod in junior high school science learning. *Jurnal Pendidikan IPA Indonesia*, 12(4), 512–523.
- Rusmini, R., Suyono, S., & Agustini, R. (2021). Development of critical thinking skills and argumentation skills assessment instruments towards 21st-century skills based on non-routine problems. *Jurnal Pendidikan Sains*, 9(2), 134–145.
- Yokhebed, Y., Karmadi, R. M. D., & Nastiti, L. R. (2025). Validity and reliability analysis of a socioscientific issues-based critical thinking self-assessment instrument using the Rasch model. *Journal of Science Education Research*, 9(1), 45–56.