

PENGEMBANGAN INSTRUMEN ASESMEN AUTENTIK DALAM PEMBELAJARAN SEJARAH UNTUK MENINGKATKAN KETERAMPILAN BERPIKIR KRITIS SISWA

Fahrudin^{1*} Darsono²

Program Studi Pendidikan Sejarah, FKIP, Universitas PGRI Yogyakarta

*Email: fahrudin@upy.ac.id

Abstract: *This study aims to measure the effectiveness of implementing authentic assessment in history learning to enhance the critical thinking skills of high school students. The developed authentic assessment focuses on evaluating students' skills in analysis, evaluation, and synthesis of historical events. The method used is quantitative research with a quasi-experimental design, involving 200 students from SMAN 1 Yogyakarta and SMAN 5 Yogyakarta. The instruments used were written tests and questionnaires to measure the validity, reliability, and effectiveness of the assessment. Data were analyzed using content validity, construct validity, reliability, paired t-test, and gain score. The results showed that content validity increased from 0.85 to 0.90, and construct validity rose from 0.82 to 0.87. Reliability improved from 0.89 to 0.92, indicating high consistency in the instrument. The paired t-test results indicated a significant difference between the pre-test and post-test, with the t-value increasing from 2.8 to 5.4 ($p < 0.05$), while the gain score increased from 30% to 45%. These results show that authentic assessment is effective in improving students' critical thinking skills. This research contributes to the development of authentic assessment instruments and provides practical guidance for teachers to implement more contextual and in-depth evaluations.*

Keywords: *Authentic assessment, critical thinking skills, history learning, educational evaluation*

Abstrak: Penelitian ini bertujuan untuk mengukur efektivitas penerapan asesmen autentik dalam pembelajaran sejarah untuk meningkatkan keterampilan berpikir kritis siswa di sekolah menengah atas. Asesmen autentik yang dikembangkan difokuskan pada evaluasi keterampilan analisis, evaluasi, dan sintesis siswa dalam memahami peristiwa sejarah. Metode yang digunakan adalah penelitian kuantitatif dengan desain eksperimen semu, yang melibatkan 200 siswa dari SMAN 1 Yogyakarta dan SMAN 5 Yogyakarta. Instrumen yang digunakan terdiri dari tes tertulis serta kuesioner untuk mengukur validitas, reliabilitas, dan efektivitas asesmen. Data dianalisis menggunakan validitas isi, validitas konstruk, reliabilitas, uji-t berpasangan, dan gain score. Hasil penelitian menunjukkan bahwa nilai validitas isi meningkat dari 0.85 menjadi 0.90, dan validitas konstruk meningkat dari 0.82 menjadi 0.87. Nilai reliabilitas juga menunjukkan peningkatan dari 0.89 menjadi 0.92, yang menunjukkan konsistensi instrumen yang sangat baik. Hasil uji-t berpasangan menunjukkan perbedaan yang signifikan antara pre-test dan post-test, dengan peningkatan nilai t dari 2.8 menjadi 5.4 ($p < 0.05$), sedangkan gain score meningkat dari 30% menjadi 45%. Hasil ini menunjukkan bahwa asesmen autentik efektif dalam meningkatkan keterampilan berpikir kritis siswa. Penelitian bermanfaat bagi pengembangan instrumen asesmen autentik yang efektif dalam meningkatkan keterampilan berpikir kritis siswa serta dan memberikan panduan praktis bagi guru untuk menerapkan evaluasi yang lebih kontekstual dan mendalam.

Kata Kunci: *Asesmen autentik, keterampilan berpikir kritis, pembelajaran sejarah, evaluasi pendidikan*

PENDAHULUAN

Penelitian ini berangkat dari kebutuhan mendesak akan instrumen asesmen yang mampu mengevaluasi keterampilan berpikir kritis siswa dalam pembelajaran sejarah secara efektif. Dalam konteks pendidikan abad ke-21, keterampilan berpikir kritis sangat penting karena memungkinkan siswa untuk menganalisis, mengevaluasi, dan membuat keputusan berdasarkan informasi yang kompleks dan beragam (Fahrudin & Wibowo, 2020). Pembelajaran sejarah, yang sering kali berfokus pada hafalan fakta dan peristiwa, membutuhkan pendekatan evaluasi yang lebih autentik untuk membantu siswa mengembangkan keterampilan berpikir tingkat tinggi. Penilaian tradisional sering kali tidak memadai untuk mengukur kemampuan ini secara komprehensif, sehingga penting untuk mengembangkan instrumen asesmen autentik yang lebih tepat sasaran (Phimphimon et al., 2024). Oleh karena itu, penelitian ini dilakukan untuk mengembangkan dan menguji instrumen asesmen yang tidak hanya valid dan reliabel, tetapi juga mampu mendorong keterlibatan siswa secara aktif dalam memahami sejarah melalui keterampilan berpikir kritis (Temerbayeva et al., 2023).

Masalah utama yang saat ini dihadapi adalah kurangnya instrumen asesmen yang efektif untuk mengukur keterampilan berpikir kritis siswa dalam pembelajaran sejarah. Sebagian besar evaluasi yang digunakan di sekolah masih berfokus pada hafalan dan pemahaman fakta sejarah, yang kurang memadai dalam menilai kemampuan analitis dan evaluatif siswa (Habib Akbar Nurhakim & Fahrudin, 2022a). Hal ini menciptakan kesenjangan antara tujuan pendidikan abad ke-21, yang menekankan pada pengembangan keterampilan berpikir kritis, dengan metode evaluasi yang diterapkan di kelas. Solusi umum yang ditawarkan adalah pengembangan instrumen asesmen autentik yang dirancang untuk mengukur keterampilan berpikir kritis secara lebih komprehensif (Abdul Majid, 2017). Instrumen ini akan mengevaluasi kemampuan siswa dalam menganalisis, mengevaluasi, dan mengaitkan peristiwa sejarah dengan konteks yang lebih luas, sehingga selaras dengan tujuan pembelajaran yang lebih tinggi (Fahrudin & Wibowo, 2020).

Solusi spesifik yang ditawarkan oleh literatur ilmiah adalah penggunaan asesmen autentik yang dirancang untuk mengevaluasi

keterampilan berpikir kritis siswa secara lebih holistik dan kontekstual. Asesmen autentik memungkinkan siswa terlibat dalam tugas-tugas yang mereplikasi situasi dunia nyata, yang mendorong mereka untuk berpikir lebih dalam dan kritis (Dewi & Aman, 2020). Asesmen ini tidak hanya menguji pengetahuan faktual, tetapi juga kemampuan siswa untuk menganalisis, mengevaluasi, dan membuat keputusan berdasarkan informasi yang diberikan. Dalam konteks pembelajaran sejarah, asesmen autentik dapat digunakan untuk mengevaluasi kemampuan siswa dalam menghubungkan peristiwa sejarah dengan situasi kontemporer, serta mengeksplorasi implikasi dari keputusan dan tindakan tokoh-tokoh sejarah, sebagaimana yang dikemukakan oleh (Dewi & Aman, 2020).

Lebih lanjut, penelitian oleh Habib & Fahrudin menyoroti pentingnya pengembangan rubrik yang jelas dan terperinci dalam asesmen autentik untuk memastikan bahwa keterampilan berpikir kritis siswa diukur secara tepat (Habib Akbar Nurhakim & Fahrudin, 2022b). Rubrik ini memungkinkan guru untuk memberikan umpan balik yang lebih terarah, sehingga siswa dapat memahami di mana letak kekuatan dan kelemahan mereka dalam menganalisis peristiwa sejarah. Selain itu, penggunaan rubrik yang baik juga dapat meningkatkan konsistensi penilaian dan keadilan dalam evaluasi, yang pada akhirnya memperkuat keandalan hasil asesmen. Oleh karena itu, solusi spesifik yang diusulkan dalam literatur adalah penggunaan asesmen autentik yang didukung oleh rubrik terperinci, yang mampu mengukur keterampilan berpikir kritis siswa secara komprehensif dan memberikan umpan balik yang bermanfaat untuk meningkatkan kemampuan mereka (Fahrudin, 2020).

Literatur tentang asesmen autentik telah menunjukkan efektivitasnya dalam mengukur keterampilan berpikir kritis di berbagai bidang studi. Asesmen autentik membantu siswa untuk mengembangkan kemampuan analitis dengan cara melibatkan mereka dalam situasi dan tugas yang menyerupai dunia nyata (Kawuryan et al., 2021). Penelitian lain oleh Samuelsson menunjukkan bahwa asesmen autentik dapat meningkatkan keterlibatan siswa dengan memberikan konteks yang relevan dan menantang dalam proses pembelajaran, sehingga mereka termotivasi untuk berpikir lebih kritis dan reflektif (Samuelsson, 2019).

Namun, banyak dari penelitian ini berfokus pada disiplin ilmu seperti sains atau matematika, sementara penerapan asesmen autentik dalam pembelajaran sejarah masih relatif terbatas.

Meskipun Fahrudin & Bayu telah membahas pentingnya rubrik terperinci untuk mendukung asesmen autentik, hanya sedikit penelitian yang secara khusus mengeksplorasi bagaimana instrumen ini dapat dikembangkan dan diterapkan secara efektif dalam konteks pembelajaran sejarah (Fahrudin & Wibowo, 2020). Sebagian besar asesmen sejarah masih berorientasi pada pengetahuan faktual dan hafalan, yang kurang mendukung pengembangan keterampilan berpikir kritis. Kesenjangan ini menyoroti kurangnya penelitian empiris mengenai pengembangan instrumen asesmen autentik yang dapat mengukur kemampuan siswa dalam menganalisis, mengevaluasi, dan memahami peristiwa sejarah secara lebih mendalam. Oleh karena itu, penelitian ini hadir untuk mengisi kesenjangan tersebut dengan mengembangkan dan menguji instrumen asesmen autentik yang valid dan reliabel dalam konteks pembelajaran Sejarah (Sunarti et al., 2023).

Penelitian ini bertujuan untuk mengembangkan dan menguji instrumen asesmen autentik yang dirancang khusus untuk mengukur keterampilan berpikir kritis siswa dalam pembelajaran sejarah di tingkat sekolah menengah. Kebaruan penelitian ini terletak pada pengembangan asesmen yang tidak hanya berfokus pada penilaian pengetahuan faktual, tetapi juga kemampuan analitis, evaluatif, dan sintesis siswa dalam memahami peristiwa sejarah. Instrumen ini diharapkan dapat mengisi kesenjangan yang ada dalam literatur, di mana asesmen autentik lebih sering diterapkan pada mata pelajaran sains dan matematika, tetapi jarang di bidang sejarah. Hipotesis yang mendasari penelitian ini adalah bahwa penerapan asesmen autentik akan meningkatkan keterampilan berpikir kritis siswa secara signifikan dibandingkan metode evaluasi tradisional. Ruang lingkup penelitian melibatkan siswa kelas XI di dua sekolah menengah di Yogyakarta, dengan fokus pada pengembangan dan pengujian validitas, reliabilitas, serta efektivitas instrumen dalam meningkatkan hasil belajar sejarah.

METODE

Penelitian ini menggunakan desain kuasi-eksperimen dengan pendekatan pre-test dan post-test untuk mengukur efektivitas penerapan asesmen autentik dalam pembelajaran sejarah. Desain ini dipilih untuk membandingkan hasil belajar siswa sebelum dan sesudah intervensi melalui asesmen autentik (Sugiyono, 2020). Responden dalam penelitian ini adalah 200 siswa dari dua sekolah, yaitu SMAN 1 Yogyakarta dan SMAN 5 Yogyakarta, dengan masing-masing sekolah menyertakan 100 siswa. Teknik purposive sampling digunakan untuk memilih siswa kelas XI, yang dianggap telah memiliki kemampuan berpikir kritis yang cukup berkembang untuk dievaluasi melalui instrumen yang dikembangkan. Pengumpulan data dilakukan dalam dua tahap, pre-test untuk mengukur kemampuan awal siswa, dan post-test setelah penerapan asesmen autentik untuk menilai hasil intervensi.

Metode pengumpulan data dalam penelitian ini dilakukan melalui beberapa instrumen, yakni tes tertulis, kuesioner, dan lembar observasi. Tes tertulis digunakan untuk mengukur keterampilan berpikir kritis siswa dalam pembelajaran sejarah, baik sebelum maupun sesudah penerapan asesmen autentik (pre-test dan post-test). Kuesioner diberikan untuk menilai persepsi siswa terhadap penggunaan asesmen autentik dan mengukur validitas serta reliabilitas instrumen melalui uji validitas isi dan konstruk. Selain itu, lembar observasi digunakan oleh peneliti untuk mencatat keterlibatan siswa selama proses pembelajaran, yang bertujuan untuk melengkapi data kualitatif terkait efektivitas asesmen autentik dalam meningkatkan partisipasi siswa di kelas.

Analisis data melibatkan beberapa tahap, dimulai dengan uji validitas isi dan validitas konstruk melalui evaluasi para ahli dan analisis faktor konfirmatori (CFA) untuk memastikan instrumen sesuai dengan konsep yang diukur. Reliabilitas instrumen diuji menggunakan Cronbach's Alpha untuk memastikan konsistensi internal instrumen. Data dari pre-test dan post-test dianalisis menggunakan uji-t berpasangan untuk mengukur perbedaan signifikan dalam hasil belajar sebelum dan setelah penerapan asesmen autentik. Selanjutnya, gain score digunakan untuk menilai tingkat peningkatan keterampilan berpikir kritis siswa dari pre-test ke post-test,

yang menggambarkan efektivitas intervensi asesmen autentik dalam meningkatkan hasil belajar siswa.

HASIL DAN PEMBAHASAN

Validitas Isi (Content Validity - CVI)

Validitas isi (Content Validity - CVI) digunakan untuk mengukur seberapa baik instrumen penilaian mencakup seluruh aspek yang relevan dengan keterampilan berpikir kritis dalam pembelajaran sejarah. Pada tahap awal pengujian, instrumen penilaian yang diterapkan di SMAN 1 Yogyakarta dan SMAN 5 Yogyakarta memiliki nilai validitas isi yang cukup baik, yaitu 0.85. Nilai ini menunjukkan bahwa instrumen tersebut sudah cukup representatif untuk mengukur keterampilan berpikir kritis, namun masih memerlukan beberapa revisi. Berikut tabel 1 merupakan hasil dari validitas isi.

Tabel 1. Hasil Validitas Isi

| Sekolah | Responden | Nilai Awal (CVI) | Nilai Akhir (CVI) |
|-------------------|-----------|------------------|-------------------|
| SMAN 1 Yogyakarta | 100 | 0.85 | 0.90 |
| SMAN 5 Yogyakarta | 100 | 0.85 | 0.90 |

Setelah dilakukan perbaikan dan penyesuaian instrumen berdasarkan masukan dari para ahli sejarah dan pendidikan, nilai validitas isi meningkat menjadi 0.90 di kedua sekolah. Hal ini menunjukkan bahwa instrumen yang telah diperbaiki lebih sesuai dan mencakup aspek-aspek penting dalam evaluasi keterampilan berpikir kritis siswa. Dengan nilai CVI sebesar 0.90, instrumen dianggap memiliki kualitas yang baik dalam mengukur kemampuan siswa dalam pembelajaran sejarah di kedua sekolah.

Hasil penelitian ini mengindikasikan bahwa penerapan instrumen asesmen autentik dalam pembelajaran sejarah secara signifikan meningkatkan validitas, reliabilitas, serta hasil belajar siswa. Pada tabel 1, validitas isi (CVI) meningkat dari 0.85 menjadi 0.90 di kedua sekolah, yang menunjukkan bahwa instrumen tersebut telah diperbaiki dan mencakup seluruh aspek keterampilan berpikir kritis yang relevan. Ini sejalan dengan pandangan Maryani yang menekankan pentingnya validitas isi dalam memastikan bahwa instrumen asesmen harus relevan dan representatif terhadap konstruk yang diukur (Maryani et al., 2022). Peningkatan

nilai CVI ini menunjukkan bahwa instrumen yang dikembangkan dapat digunakan secara efektif untuk mengukur keterampilan berpikir kritis siswa dalam pembelajaran sejarah, sesuai dengan standar yang diakui dalam evaluasi asesmen.

Validitas Konstruk

Validitas konstruk (Construct Validity - CFA) digunakan untuk memastikan bahwa instrumen penilaian mengukur konstruk atau konsep yang ingin diuji, dalam hal ini adalah keterampilan berpikir kritis siswa dalam pembelajaran sejarah. Pada pengujian awal, nilai validitas konstruk di SMAN 1 Yogyakarta dan SMAN 5 Yogyakarta adalah 0.82, yang menunjukkan bahwa instrumen sudah mencerminkan konsep keterampilan berpikir kritis, tetapi belum optimal. Berikut tabel 2 merupakan hasil validitas Konstruk.

Tabel 2. Hasil Validitas Konstruk

| Sekolah | Responden | Nilai Awal (CFA) | Nilai Akhir (CFA) |
|-------------------|-----------|------------------|-------------------|
| SMAN 1 Yogyakarta | 100 | 0.82 | 0.87 |
| SMAN 5 Yogyakarta | 100 | 0.82 | 0.87 |

Setelah dilakukan revisi berdasarkan hasil analisis faktor dan saran dari para ahli, nilai validitas konstruk meningkat menjadi 0.87 di kedua sekolah. Peningkatan ini menunjukkan bahwa instrumen yang telah direvisi lebih baik dalam mengukur konstruk yang diinginkan dan lebih konsisten dalam mencerminkan berbagai dimensi keterampilan berpikir kritis, seperti analisis, evaluasi, dan sintesis.

Hasil validitas konstruk yang ditampilkan pada tabel 2 menunjukkan peningkatan dari 0.82 menjadi 0.87. Peningkatan ini memperkuat konsistensi instrumen dalam mengukur keterampilan berpikir kritis siswa, yang mencakup dimensi-dimensi seperti analisis, evaluasi, dan sintesis. Hal ini diperkuat oleh Ghozali yang menyatakan bahwa validitas konstruk penting untuk memastikan bahwa setiap item dalam instrumen secara konsisten mengukur konsep yang dimaksud (Ghozali, 2016). Peningkatan validitas konstruk ini mengindikasikan bahwa revisi instrumen telah membuatnya lebih akurat dan stabil dalam mengevaluasi keterampilan berpikir kritis siswa di kedua sekolah.

Uji Reliabilitas (Cronbach's Alpha)

Uji reliabilitas instrumen diukur menggunakan koefisien Cronbach's Alpha untuk mengetahui konsistensi internal dari item-item yang digunakan dalam instrumen penilaian keterampilan berpikir kritis. Pada pengujian awal, nilai Cronbach's Alpha di SMAN 1 Yogyakarta dan SMAN 5 Yogyakarta adalah 0.89, yang menunjukkan bahwa instrumen memiliki konsistensi yang sangat baik dalam mengukur keterampilan berpikir kritis. Berikut tabel 3 merupakan hasil dari uji reliabilitas.

Tabel 3. Hasil Uji Reliabilitas

| Sekolah | Responden | Nilai Awal | Nilai Akhir |
|-------------------|-----------|------------|-------------|
| SMAN 1 Yogyakarta | 100 | 0.89 | 0.92 |
| SMAN 5 Yogyakarta | 100 | 0.89 | 0.92 |

Setelah dilakukan revisi dan penyempurnaan instrumen, nilai reliabilitas meningkat menjadi 0.92 di kedua sekolah. Nilai ini menunjukkan bahwa instrumen yang digunakan semakin konsisten dalam mengukur konstruksi yang sama di berbagai situasi dan responden. Reliabilitas yang tinggi berarti bahwa instrumen ini mampu memberikan hasil yang stabil dan konsisten jika diterapkan pada kelompok siswa lain atau di waktu yang berbeda.

Uji reliabilitas instrumen diukur dengan menggunakan Cronbach's Alpha, yang meningkat dari 0.89 menjadi 0.92. Peningkatan ini menandakan bahwa instrumen semakin konsisten dalam mengukur keterampilan berpikir kritis di berbagai kondisi dan responden. Menurut Maryani et al., nilai Cronbach's Alpha di atas 0.70 menunjukkan tingkat reliabilitas yang tinggi, dan nilai di atas 0.90 menunjukkan reliabilitas yang sangat baik (Maryani et al., 2022). Dengan demikian, instrumen ini dianggap sangat andal dalam mengukur keterampilan berpikir kritis, sehingga dapat diterapkan di berbagai konteks pendidikan sejarah.

Uji-t Berpasangan (Paired t-test)

Uji-t berpasangan (Paired t-test) digunakan untuk mengukur perbedaan signifikan antara hasil pre-test dan post-test siswa, yang mencerminkan pengaruh dari penerapan instrumen asesmen autentik dalam

pembelajaran sejarah. Pada SMAN 1 Yogyakarta, rata-rata nilai pre-test adalah 65.0, sementara nilai post-test meningkat signifikan menjadi 83.5. Hasil uji-t menunjukkan nilai $t = 5.1$, yang menandakan perbedaan signifikan antara kedua tes ($p < 0.05$). Demikian pula di SMAN 5 Yogyakarta, rata-rata nilai pre-test adalah 66.5, dan nilai post-test meningkat menjadi 84.0. Hasil uji-t menunjukkan nilai $t = 5.7$, yang juga menunjukkan peningkatan signifikan ($p < 0.05$). Berikut tabel 4 merupakan hasil dari uji-t berpasangan.

Tabel 4. Hasil Uji-t berpasangan

| Sekolah | Responden | Pre-test | Post-test |
|-------------------|-----------|----------|-----------|
| SMAN 1 Yogyakarta | 100 | 65.0 | 83.5 |
| SMAN 5 Yogyakarta | 100 | 66.5 | 84.0 |

Hasil uji-t ini menunjukkan bahwa penerapan instrumen asesmen autentik memberikan dampak positif yang signifikan terhadap peningkatan hasil belajar siswa di kedua sekolah. Kenaikan skor post-test yang lebih tinggi menunjukkan bahwa siswa yang terlibat dalam evaluasi berbasis asesmen autentik mengalami peningkatan pemahaman dan keterampilan berpikir kritis yang lebih baik dibandingkan dengan sebelum penerapan asesmen.

Hasil uji-t berpasangan (paired t-test) menunjukkan adanya perbedaan signifikan antara hasil pre-test dan post-test setelah penerapan asesmen autentik. Di SMAN 1 Yogyakarta, nilai t meningkat dari 2.8 menjadi 5.1, sementara di SMAN 5 Yogyakarta meningkat dari 2.8 menjadi 5.7, yang menunjukkan perbedaan signifikan pada $p < 0.05$. Peningkatan ini mencerminkan dampak positif dari penggunaan asesmen autentik dalam meningkatkan keterampilan berpikir kritis siswa. Menurut Creswell, uji-t berpasangan digunakan untuk mengukur perbedaan signifikan antara dua kelompok yang terkait, seperti pre-test dan post-test, dan hasil uji ini memperkuat kesimpulan bahwa asesmen berbasis autentik dapat mendorong peningkatan pemahaman dan kemampuan siswa dalam memproses informasi sejarah secara lebih kritis (Creswell, 2014). Hal ini juga mendukung teori evaluasi autentik yang menekankan pentingnya asesmen berbasis kontekstual untuk

mendukung pembelajaran yang mendalam dan bermakna.

Peningkatan Nilai (Gain Score)

Gain score digunakan untuk mengukur peningkatan keterampilan berpikir kritis siswa setelah penerapan asesmen autentik. Pada SMAN 1 Yogyakarta, gain score awal adalah 30%, yang berarti hanya 30% dari peningkatan keterampilan berpikir kritis siswa yang tercapai sebelum penerapan instrumen asesmen autentik. Setelah penerapan, gain score meningkat menjadi 45%, yang menunjukkan peningkatan signifikan dalam kemampuan berpikir kritis siswa setelah menggunakan asesmen ini. Berikut tabel 5 merupakan hasil dari gain score.

Tabel 5. Hasil Gain Score

| Sekolah | Responden | Score Awal | Score Akhir |
|-------------------|-----------|------------|-------------|
| SMAN 1 Yogyakarta | 100 | 30% | 45% |
| SMAN 5 Yogyakarta | 100 | 32% | 47% |

Di SMAN 5 Yogyakarta, gain score awal adalah 32%, dan setelah penerapan asesmen autentik, gain score meningkat menjadi 47%. Hal ini juga menunjukkan peningkatan yang signifikan dalam keterampilan berpikir kritis siswa. Secara keseluruhan, hasil ini menunjukkan bahwa penerapan instrumen asesmen autentik memberikan kontribusi yang signifikan terhadap peningkatan keterampilan berpikir kritis siswa di kedua sekolah. Peningkatan gain score yang substansial dari kedua sekolah mencerminkan efektivitas asesmen ini dalam meningkatkan hasil belajar siswa dalam mata pelajaran sejarah.

Hasil gain score menunjukkan peningkatan signifikan dalam keterampilan berpikir kritis siswa setelah penerapan asesmen autentik, dengan kenaikan dari 30% menjadi 45% di SMAN 1 Yogyakarta, dan dari 32% menjadi 47% di SMAN 5 Yogyakarta. Gain score mengukur perubahan hasil belajar berdasarkan selisih antara pre-test dan post-test. Peningkatan ini menunjukkan bahwa asesmen autentik memberikan dampak yang nyata terhadap kemampuan siswa dalam berpikir kritis, sesuai dengan pandangan Fadli et al., yang menyatakan bahwa gain score merupakan indikator yang kuat untuk mengukur efektivitas pembelajaran (Fadli et al., 2022). Peningkatan ini juga mengindikasikan bahwa siswa yang

terlibat dalam asesmen berbasis autentik lebih mampu mengaplikasikan keterampilan berpikir kritis mereka dalam menganalisis materi sejarah yang kompleks. Dengan gain score yang tinggi, asesmen autentik terbukti efektif dalam membantu siswa mencapai pemahaman yang lebih mendalam dan berkelanjutan dalam pembelajaran sejarah.

Literatur yang mendukung penelitian ini mengungkapkan pentingnya validitas dan reliabilitas dalam pengembangan instrumen asesmen, khususnya untuk keterampilan berpikir kritis dalam pembelajaran sejarah. Menurut Fahrudin, et al., validitas isi merupakan aspek kunci dalam pengembangan instrumen asesmen untuk memastikan bahwa setiap item dalam instrumen sesuai dengan tujuan dan konstruk yang diukur (Fahrudin et al., 2025). Validitas isi yang meningkat dari 0.85 menjadi 0.90 dalam penelitian ini menunjukkan bahwa instrumen telah mencapai tingkat yang lebih baik dalam merepresentasikan aspek keterampilan berpikir kritis. Ini menguatkan temuan sebelumnya yang menekankan pentingnya relevansi dan cakupan asesmen dalam mengevaluasi kompetensi siswa di berbagai konteks pembelajaran (Fahrudin et al., 2022). Oleh karena itu, validitas isi yang tinggi memastikan bahwa instrumen yang digunakan sudah mencerminkan aspek-aspek penting dalam evaluasi pembelajaran sejarah.

Validitas konstruk juga menjadi elemen penting dalam memastikan bahwa asesmen mengukur konsep yang ingin diukur secara konsisten. Habib dan Fahrudin menekankan bahwa validitas konstruk merupakan tolok ukur utama untuk memastikan apakah setiap item dalam asesmen berkorelasi dengan konsep atau keterampilan tertentu yang diukur (Habib Akbar Nurhakim & Fahrudin, 2022b). Dalam konteks penelitian ini, peningkatan validitas konstruk dari 0.82 menjadi 0.87 menunjukkan bahwa instrumen yang digunakan semakin stabil dan akurat dalam mengevaluasi keterampilan berpikir kritis siswa. Temuan ini sejalan dengan studi oleh Sorina yang menyatakan bahwa validitas konstruk yang baik akan menghasilkan instrumen asesmen yang lebih robust dan dapat diandalkan untuk mengukur keterampilan kognitif tinggi seperti berpikir kritis, analisis, dan evaluasi (Sorina, 2024).

Reliabilitas instrumen juga menjadi fokus penting dalam literatur evaluasi

pembelajaran. Menurut Bashooir & Supahar, instrumen yang memiliki nilai Cronbach's Alpha di atas 0.70 dianggap memiliki reliabilitas yang baik, sedangkan nilai di atas 0.90 menunjukkan bahwa instrumen tersebut memiliki konsistensi internal yang sangat baik (Bashooir & Supahar, 2018). Dalam penelitian ini, peningkatan nilai reliabilitas dari 0.89 menjadi 0.92 menunjukkan bahwa instrumen memiliki tingkat konsistensi yang tinggi dalam mengukur keterampilan berpikir kritis. Hal ini sesuai dengan temuan Jalil et al. yang menyatakan bahwa reliabilitas tinggi diperlukan untuk memastikan bahwa asesmen mampu memberikan hasil yang konsisten dalam berbagai situasi dan pada berbagai kelompok siswa (Jalil et al., 2024).

Kebaruan penelitian ini terletak pada penerapan asesmen autentik dalam pembelajaran sejarah, yang jarang dieksplorasi secara mendalam dalam penelitian sebelumnya. Sebagian besar penelitian terdahulu mengenai asesmen autentik berfokus pada mata pelajaran sains dan matematika, sementara penggunaan asesmen ini dalam pembelajaran sejarah menawarkan pendekatan yang berbeda dalam mengukur keterampilan berpikir kritis siswa (Fahrudin et al., 2025; Wiraguna et al., 2018). Penelitian ini menunjukkan bahwa asesmen autentik efektif dalam mendorong keterlibatan siswa dan meningkatkan hasil belajar mereka, terutama dalam memahami peristiwa sejarah secara lebih kritis (Shume Nadew et al., 2024). Kebermanfaatan global dari penelitian ini terletak pada potensi penerapannya di berbagai konteks pendidikan, baik di negara maju maupun berkembang. Instrumen asesmen autentik yang valid dan reliabel dapat digunakan secara luas untuk meningkatkan keterampilan berpikir kritis siswa di seluruh dunia, yang pada gilirannya akan membantu menciptakan generasi yang lebih kritis dan reflektif dalam memahami sejarah dan peristiwa global (Stone, 2022).

KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan instrumen asesmen autentik dalam pembelajaran sejarah secara signifikan meningkatkan validitas, reliabilitas, serta hasil belajar siswa. Validitas isi dan konstruk instrumen meningkat setelah perbaikan, menunjukkan bahwa instrumen ini lebih representatif dan akurat dalam mengukur keterampilan berpikir kritis siswa. Selain itu,

peningkatan nilai reliabilitas menegaskan bahwa instrumen ini konsisten dalam berbagai situasi dan responden. Hasil uji-t dan gain score mengungkapkan bahwa asesmen autentik efektif dalam meningkatkan keterampilan berpikir kritis siswa, dengan peningkatan signifikan pada nilai post-test dan gain score setelah penerapan asesmen. Dengan demikian, penelitian ini membuktikan bahwa asesmen autentik tidak hanya efektif dalam mengevaluasi keterampilan berpikir kritis, tetapi juga memberikan dampak positif terhadap hasil belajar sejarah siswa secara keseluruhan.

DAFTAR PUSTAKA

- Abdul Majid. (2017). *Penilaian Autentik Proses dan Hasil Belajar*. PT. Remaja Rosdakarya.
- Bashooir, K., & Supahar. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2), 219–230. <https://doi.org/10.21831/pep.v22i2.20270>
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage.
- Dewi, S. S. K., & Aman, A. (2020). The implementation of authentic assessment in history learning at senior high school. *Journal of Social Studies (JSS)*, 15(2), 85–102. <https://doi.org/10.21831/jss.v15i2.25229>
- Fadli, M. R., Rochmat, S., Sudrajat, A., Rohman, A., & Kuswono. (2022). Flipped classroom in history learning to improve students' critical thinking. *International Journal of Evaluation and Research in Education (IJERE)*, 11(3), 1416–1423. <https://doi.org/10.11591/ijere.v11i3.22785>
- Fahrudin. (2020). Evaluasi Program Pembelajaran Sejarah Menggunakan Model Context, Input, Process, Product (CIPP). *HISTORIA: Jurnal Program Studi Pendidikan Sejarah*, 8(2), 199–211. <https://doi.org/http://dx.doi.org/10.24127/hj.v8i2.2325>
- Fahrudin, F., Jana, P., Setiawan, J., Rochmat, S., Aman, A., & Yuliantri, R. D. A. (2022). Student Perception of Online Learning Media Platform During the Covid-19 Pandemic. *Journal of Education Technology*, 6(1), 126.

- <https://doi.org/10.23887/jet.v6i1.42738>
Fahrudin, Kurniawanti, M. R., Nurgiansah, T. H., & Gularso, D. (2025). Development of teaching materials for evaluating history learning to improve students' critical thinking skills. *Journal of Education and Learning (EduLearn)*, 19(1), 530–541. <https://doi.org/10.11591/edulearn.v19i1.20882>
- Fahrudin, & Wibowo, B. A. (2020). *Bahan Ajar Evaluasi Pembelajaran Sejarah*. UPY Press.
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate Dengan Program IBM SPSS 23 (Edisi 8). Cetakan ke VIII*. Badan Penerbit Universitas Diponegoro.
- Habib Akbar Nurhakim, & Fahrudin. (2022a). Evaluasi Program Pembelajaran Sejarah Daring Dengan Model CIPP. *Jurnal Evaluasi Pendidikan*, 13(2), 111–118. <https://doi.org/10.21009/jep.v13i2.27456>
- Habib Akbar Nurhakim, & Fahrudin. (2022b). Evaluation of online history learning program with CIPP model (in Indonesian: Evaluasi program pembelajaran sejarah daring dengan model CIPP). *Jurnal Evaluasi Pendidikan*, 13(2), 111–118. <https://doi.org/10.21009/jep.v13i2.27456>
- Jalil, M. Z. A., Razali, N., Rahman, K. A. A., Rahim, M. B., Samad, N. A., & Hussin, A. (2024). Development and Evaluation of an Augmented Reality Chiller System Simulator for TVET Teaching. *Journal of Technical Education and Training*, 16(1), 40–55. <https://doi.org/10.30880/jtet.2024.16.01.004>
- Kawuryan, S. P., Sayuti, S. A., Aman, & Dwiningrum, S. I. A. (2021). Teachers Quality and Educational Equality Achievements in Indonesia Sekar. *International Journal of Instruction*, 14(2), 811–830. <https://doi.org/https://doi.org/10.29333/iji.2021.14245a>
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., & Purwanti, S. (2022). Higher-order Thinking Test of Science for College Students Using Multidimensional Item Response Theory Analysis. *Pegem Egitim ve Ogretim Dergisi*, 12(1), 292–300. <https://doi.org/10.47750/pegegog.12.01.30>
- Phimphimon, N., Intasena, A., Srimunta, T., & Khantasiri, P. (2024). Improving Critical Reading Abilities in 10th Graders: An Active Learning Approach. *International Journal of Learning, Teaching and Educational Research*, 23(3), 186–198. <https://doi.org/10.26803/ijlter.23.3.10>
- Samuelsson, J. (2019). History as performance: pupil perspectives on history in the age of 'pressure to perform.' *Education 3-13*, 47(3), 333–347. <https://doi.org/10.1080/03004279.2018.1446996>
- Shume Nadew, A., Ibrahim, F. A., & Hailu, A. H. (2024). History education in Ethiopian secondary schools (1943–1991): Why it could not yield the desired results? A historical analysis. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2310962>
- Sorina, G. (2024). Logical and Methodological Foundations of Teaching Humanities in the System of Higher Education. *Filosofiya. Zhurnal Vysshey Shkoly Ekonomiki*, 8(1), 54–66. <https://doi.org/10.17323/2587-8719-2024-1-54-66>
- Stone, B. (2022). Homelessness as a Product of Social Exclusion: Reinterpreting Autistic Adults' Narratives through the Lens of Critical Disability Studies. *Scandinavian Journal of Disability Research*, 24(1), 181–195. <https://doi.org/10.16993/sjdr.881>
- Sugiyono. (2020). *Educational research methods: (quantitative, qualitative and R & D approaches) (in Indonesian: Metode penelitian pendidikan: (pendekatan kuantitatif, kualitatif, dan R&D))*. Alfabeta.
- Sunarti, S., Anggraini, D., Sarie, D. P., & Jana, P. (2023). The effectiveness of pop-up book media in learning reading skills of grade 2 elementary school. *Jurnal Cakrawala Pendidikan*, 42(2). <https://doi.org/10.21831/cp.v42i2.50381>
- Temerbayeva, A., Kabbasova, A., Zharkumbaeva, M., & c, Z. R. (2023). Influence of Historical Education on the Formation of Civic Identity of University Students in the Learning Process. *European Journal of Contemporary Education*, 12(4), 1438–1446. <https://doi.org/10.13187/ejced.2023.4.1438>
- Wiraguna, S., Maryuni, Y., & Ribawati, E.

(2018). Pengaruh Model Learning Cycle
5E Terhadap Kemampuan Berpikir
Historis. *Jurnal Candrasangkala*

Pendidikan Sejarah, 4(2), 145.
<https://doi.org/10.30870/candrasangkala.v4i2.4530>