

PENGEMBANGAN INSTRUMEN PENILAIAN E-HOTS DALAM MENINGKATKAN KRITIS ANALITIS MAHASISWA PROGRAM STUDI PENDIDIKAN SEJARAH

Oleh :

Ulfah Nury Batubara¹⁾, Erwin Siregar²⁾, Aulia Ramadhani³⁾, Parlagutan Nasution⁴⁾

^{1,2,3,4} Institut Pendidikan Tapanuli Selatan
email: ulfahnury@gmail.com

Informasi Artikel

Riwayat Artikel :

Submit, 30 September 2025

Revisi, 9 Januari 2026

Diterima, 14 Januari 2026

Publish, 15 Januari 2026

Kata Kunci :

Instrument,
Penilaian,
EHOTS,
Pembelajaran Sejarah.

ABSTRAK

Pembelajaran merupakan serangkaian proses yang dilakukan oleh guru agar peserta didik belajar. Proses tersebut berisi seperangkat aktivitas yang dilakukan peserta didik untuk mencapai tujuan pembelajaran. Sudjana menyatakan bahwa komponen-komponen penting dalam sebuah pembelajaran ada empat, yakni: tujuan, bahan ajar, metode, dan alat/penilaian. Kusaeri et.al menyatakan bahwa penilaian adalah suatu prosedur sistematis mulai dari mengumpulkan, menganalisis, serta menginterpretasikan informasi, yang digunakan untuk membuat kesimpulan tentang karakteristik seseorang atau objek. Melalui penilaian, seorang guru dapat mengambil keputusan apakah peserta didik sudah memenuhi target yang diinginkan atau masih memerlukan tindakan perbaikan. Standar penilaian pada abad 21 dilakukan dengan mengadaptasi model-model penilaian berstandar internasional salah satunya adalah HOTS (*Higher Order Thinking Skill*) untuk menilai apakah peserta didik sudah memiliki kemampuan berpikir tingkat tinggi. Penelitian ini bertujuan untuk mengembangkan instrument penilaian EHOTS dalam meningkatkan kritis analitis mahasiswa program studi Pendidikan Sejarah. Tempat penelitian dilakukan di prodi Pendidikan Sejarah Institut Pendidikan Tapanuli Selatan. Adapun teknik pengembangan instrumen penilaian dikembangkan melalui lima langkah yaitu (1) penelitian dan pengumpulan informasi, (2) perencanaan, (3) pengembangan produk awal, (4) uji coba terbatas, (5) revisi produk awal, yang diadopsi dari langkah pengembangan Borg & Gall. Hasil penelitian menunjukkan bahwa validitas soal EHOTS dari 25 item soal terdapat 3 soal yang tidak valid. Sementara tingkat reliabilitas soal HOTS adalah 0,5899 dengan kategori sedang atau cukup, tingkat kesukaran soal EHOTS memiliki rata-rata 0,6168 dengan kategori cukup. Selanjutnya rata-rata daya pembeda yakni 0,2728 dengan kategori cukup.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



Corresponding Author:

Nama: Ulfah Nury Batubara

Afiliasi: Institut Pendidikan Tapanuli Selatan

Email: ulfahnury@gmail.com

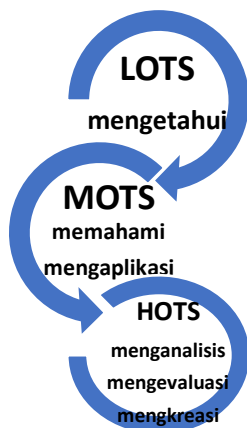
1. PENDAHULUAN

Pembelajaran merupakan serangkaian proses yang dilakukan oleh guru agar peserta didik belajar. Proses tersebut berisi seperangkat aktivitas yang dilakukan peserta didik untuk mencapai tujuan pembelajaran (Abidin, 2012:3). Sudjana menyatakan bahwa komponen-komponen penting dalam sebuah

pembelajaran ada empat, yakni: tujuan, bahan ajar, metode, dan alat/penilaian (2010: 30). Tujuan merupakan komponen utama dalam pembelajaran, karena memuat sasaran atau target yang harus dicapai oleh peserta didik serta mau dibawa ke mana arah pembelajaran semuanya tergantung pada tujuan pembelajaran (Sanjaya, 2006:59). Kedua, bahan ajar

atau materi yakni segala bahan (baik informasi, alat maupun teks) yang disusun secara sistematis, menampilkan secara utuh kompetensi yang akan dikuasai oleh peserta didik (Prastowo, 2012:17). Ketiga, metode ialah cara-cara untuk mencapai hasil pembelajaran di bawah kondisi yang berbeda. Pemilihan metode pembelajaran harus disesuaikan dengan situasi dan kondisi peserta didik agar tujuan pembelajaran tercapai.

Komponen terakhir yang paling penting dalam sebuah pembelajaran adalah penilaian. Kusaeri et.al menyatakan bahwa penilaian adalah suatu prosedur sistematis mulai dari mengumpulkan, menganalisis, serta menginterpretasikan informasi, yang digunakan untuk membuat kesimpulan tentang karakteristik seseorang atau objek (Kusaeri & Suprananto 2012:8). Melalui penilaian, seorang guru dapat mengambil keputusan apakah peserta didik sudah memenuhi target yang diinginkan atau masih memerlukan tindakan perbaikan. Standar penilaian pada abad 21 dilakukan dengan mengadaptasi model-model penilaian berstandar internasional salah satunya adalah HOTS (*Higher Order Thinking Skill*) untuk menilai apakah peserta didik sudah memiliki kemampuan berpikir tingkat tinggi. Kemampuan berpikir ini dibagi dalam tiga bidang yakni *Lower Order Thinking Skill*, *Medium Order Thinking Skill* serta *Higher Order Thinking Skill*.



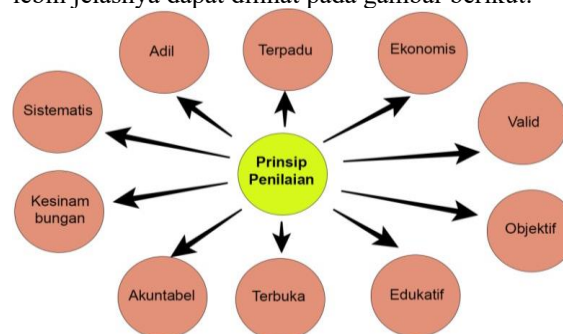
Gambar 1 : Dimensi Proses Berpikir

Penilaian yang berstandar Internasional tersebut juga harus diterapkan dalam pembelajaran sejarah. Pembelajaran sejarah berkenaan dengan upaya memperkenalkan peserta didik terhadap ilmu sejarah. Oleh karena itu, kemampuan berpikir kritis seperti berpikir kronologis, kemampuan analisis dalam penafsiran sejarah, serta pengambilan keputusan menjadi tujuan penting dalam pembelajaran sejarah. Sejarah bukan sekadar fakta atau peristiwa yang terjadi, tetapi juga proses yang melibatkan berbagai faktor sosial, politik, ekonomi, dan budaya. Melalui berpikir kritis, seseorang yang belajar sejarah dapat menggali konteks dan memahami mengapa suatu peristiwa terjadi dan dampaknya untuk hari ini. Berfikir analitis membantu untuk memisahkan fakta dan interpretasi untuk kepentingan tertentu. Selain itu, berpikir kritis dapat

mengevaluasi kebenaran informasi yang disajikan dan mempertanyakan sumber dan tujuan informasi tersebut. Sejarah memberi pelajaran penting tentang pola atau tren yang dapat membantu memahami masa depan (Mardapi 1999: 45). Dengan berpikir kritis, dapat mengidentifikasi pola yang mungkin muncul kembali, seperti konflik atau perubahan sosial, dan belajar dari pengalaman masa lalu untuk menghindari kesalahan yang sama di masa depan.

Berdasarkan hasil observasi yang dilakukan peneliti terhadap mahasiswa program studi pendidikan sejarah angkatan 2023 Institut Pendidikan Tapanuli Selatan pada 26 Februari 2025 diketahui bahwa mahasiswa masih belum mampu untuk berfikir kritis analitis. Banyak mahasiswa masih sekedar menghafal tahun dan peristiwa sejarah. Ketika mempelajari suatu peristiwa sejarah tidak berfikir kausalitas atau sebab akibat, dari suatu peristiwa. Hal ini bisa juga terjadi sebab mahasiswa tidak dibiasakan untuk mengevaluasi peristiwa sejarah. Sejalan dengan itu hasil penelitian oleh Windia Hadi, et all yang menyatakan bahwa faktor yang menyebabkan mahasiswa masih lemah dalam menyelesaikan masalah bermuatan HOTS adalah ketidakbiasaan dan kurang dilatih untuk memiliki kemampuan berfikir tingkat tinggi dalam menyelesaikan masalah.

Ketidakmampuan mahasiswa dalam berfikir tingkat juga terdapat dalam buku panduan Kemenristekdikti yang menyatakan bahwa meskipun pendidikan di Indonesia sudah berusaha untuk mengembangkan keterampilan berpikir tingkat tinggi, banyak mahasiswa yang masih bergantung pada ingatan dan pemahaman dasar daripada menerapkan analisis, sintesis, atau evaluasi dalam menyelesaikan masalah yang kompleks. Hal ini menjadi alasan dasar bagi peneliti untuk mengembangkan instrumen penilaian EHOTS (*Evaluation of Higher order thinking Skill*). Prinsip-prinsip dalam penilaian yang akan dikembangkan lebih jelasnya dapat dilihat pada gambar berikut:



Gambar 2 : Bagan prinsip penilaian

Dalam penyusunan instrumen penilaian EHOTS, kreativitas seorang pendidik sangat mempengaruhi kualitas dan variasi stimulus yang digunakan. Adapun rumusan masalah dalam penelitian ini yakni: bagaimana pengembangan instrumen penilaian EHOTS (*evaluation of Higher order thinking Skill*) dalam meningkatkan kritis

analitis mahasiswa program studi pendidikan sejarah?

2. METODE PENELITIAN

Penelitian ini merupakan penelitian pengembangan yang mengacu pada jenis penelitian *Research and Development* (R & D). Model pengembangan ini dipandang cocok karena mengembangkan instrumen penilaian kemampuan berpikir tingkat tinggi (*Higher Order Thinking Skill*) mahasiswa prodi pendidikan Sejarah. Prosedur pengembangan yang digunakan dalam penelitian ini menggunakan model pengembangan dari Borg & Gall (2003: 772). Penelitian pengembangan menurut Borg & Gall adalah penelitian yang berorientasi untuk mengembangkan dan memvalidasi produk-produk yang digunakan dalam pendidikan. Produk yang ingin dihasilkan dalam penelitian ini berupa instrumen penilaian EHOTS khususnya materi perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda. Rangkaian tahapan penelitian pengembangan yang dikemukakan oleh Borg & Gall terdiri dari 10 (sepuluh) tahapan, yaitu: 1) *research and information collecting*, 2) *planning*, 3) *develop preliminary form of product*, 4) *preliminary field testing*, 5) *main product revision*, 6) *main field testing*, 7) *operasional product revision*, 8) *operasional field testing*, 9) *final product revision*, dan 10) *dissemination and implementation*. Tempat penelitian dilaksanakan pada mahasiswa Pendidikan Sejarah Institut Pendidikan Tapanuli Selatan yang melibatkan 30 mahasiswa. Selanjutnya soal yang akan dikembangkan akan diuji validitas, reliabilitas, tingkat kesukaran dan daya pembeda.

3. HASIL DAN PEMBAHASAN

a. Konsep Dasar Penilaian

Definisi penilaian menurut Arifin (2009: 2) yakni sebuah proses atau kegiatan sistematis dan berkesinambungan untuk mengumpulkan informasi tentang proses dan hasil belajar peserta didik dalam rangka membuat keputusan berdasarkan kriteria dan pertimbangan tertentu. Sementara itu, Ridwan (2016:15) mengungkapkan bahwa penilaian adalah upaya sistematis dan sistemik yang dilakukan melalui pengumpulan data atau informasi yang shahih (valid) dan reliabel, selanjutnya data atau informasi tersebut diolah sebagai upaya melakukan pertimbangan untuk pengambilan kebijakan suatu program pendidikan. Tidak jauh berbeda dengan pendapat di atas, penilaian menurut Kusaeri & Suprananto (2012:8) adalah suatu prosedur sistematis

dan mencakup kegiatan mengumpulkan, menganalisis, serta menginterpretasikan informasi yang dapat digunakan untuk membuat kesimpulan tentang karakteristik seseorang atau objek. Dalam dunia pendidikan informasi tersebut untuk menentukan seberapa jauh peserta didik mencapai tujuan pembelajaran yang telah ditetapkan, baik aspek pengetahuan, sikap maupun keterampilan.

Penilaian juga dapat diartikan sebagai kegiatan menafsirkan atau memaknai data hasil suatu pengukuran berdasarkan kriteria atau standar tertentu. Dengan kata lain penilaian sebagai pemberian makna atau ketetapan kualitas hasil suatu pengukuran dengan cara membandingkan data hasil pengukuran (Eko. 2014: 4). Penilaian yang baik dan teliti akan memberikan deskripsi proses dan output hasil belajar yang objektif. Oleh karena itu, sistem penilaian yang digunakan dalam lembaga pendidikan harus bisa: (1) memberikan informasi yang akurat; (2) mendorong peserta didik belajar; (3) memotivasi tenaga pendidik mengajar; (4) meningkatkan kinerja lembaga, dan (5) meningkatkan kualitas pendidikan (Kate, 2017: 110). Penilaian merupakan bagian terpenting dari sebuah pembelajaran, sebab penilaian berisi sejumlah fakta yang dapat menjelaskan karakteristik seseorang atau objek yang dinilai. Dalam dunia pendidikan, perkembangan kemampuan peserta didik tidak akan dapat diketahui tanpa penilaian (Wahyono, 2017: 20). Penilaian yang baik akan memberi dampak pada proses pembelajaran dan menjadi rujukan untuk kebijakan selanjutnya. Oleh karena itu, pemilihan metode penilaian akan berpengaruh terhadap validitas hasil penilaian yang berisi informasi objektif dan valid. Sebaliknya kesalahan dalam memilih metode penilaian akan berimbas pada informasi yang tidak valid dan tidak objektif sesuai dengan data di lapangan (Setiadi, 2016: 167).

Dalam melakukan penilaian seorang guru harus paham ciri-ciri dari penilaian. Adapun ciri-ciri penilaian sebagaimana dirumuskan oleh Arikunto (2002: 11) ada lima, yaitu: (1) penilaian dapat dilakukan secara tidak langsung, contoh mengetahui tingkat inteligen seorang anak diukur dengan kemampuan menyelesaikan soal; (2) penilaian bersifat kuantitatif artinya menggunakan simbol bilangan sebagai hasil pertama pengukuran. Setelah itu, diinterpretasikan dalam bentuk kualitatif. Sementara ciri ketiga (3) dari penilaian adalah menggunakan satuan-satuan atau unit yang tetap. Ciri keempat (4) bahwa penilaian bersifat relatif artinya tidak sama atau tidak selalu tetap dari satu waktu ke waktu yang lain. Terakhir ciri kelima (5) adalah penilaian sering terjadi kesalahan-kesalahan, sehingga seorang guru harus lebih teliti dalam melakukan penilaian, baik metode yang digunakan maupun sewaktu melakukan penilaian terhadap objek. Penilaian dinilai efektif jika dilakukan secara berkesinambungan dan seimbang (Chiapetta & Koballa, 2010: 35). Berkesinambungan yang

dimaksud adalah penilaian dilakukan sepanjang pembelajaran berlangsung. Penilaian juga dituntut seimbang, yakni melibatkan gabungan dari teknik penilaian alternatif dan tradisional. Penilaian dapat dilakukan sepanjang pembelajaran melalui berbagai cara yang beragam (Abolfazli & Sadeghi, 2013: 1553). Selain menjadi bagian dari kegiatan belajar-mengajar, penilaian juga dapat menentukan kesuksesan suatu pembelajaran (Osadebe, 2014: 16). Meski demikian penilaian tidak hanya mengukur proses dan keluaran pembelajaran saja, tetapi juga dapat mengembangkan proses pembelajaran, memotivasi peserta didik dan mendiagnosa kesulitan peserta didik (Fatonah & Prasetyo, 2013: 53). Penilaian juga dipandang sebagai kegiatan pengumpulan bukti-bukti peserta didik yang berkaitan dengan proses pembelajaran sebagai informasi untuk mengambil keputusan dalam kegiatan belajar mengajar selanjutnya (Stiggins & Chappius, 2014: 3). Guru menggunakan penilaian untuk menentukan tingkat kemampuan peserta didik melalui interpretasi tugas rutin peserta didik. Oleh karena itu, guru dapat memberikan umpan balik terhadap capaian peserta didik dalam proses pembelajaran (Webb, 2012: 60).

Berdasarkan penjelasan beberapa ahli tersebut, maka dapat ditarik kesimpulan bahwa penilaian adalah proses untuk mengetahui kemampuan dan tingkat keberhasilan peserta didik dalam belajar. Kegiatan ini menggunakan metode pengukuran dan dilakukan secara menyeluruh dan berkesinambungan sehingga dapat dijadikan umpan balik dalam pembelajaran. Penilaian dalam pembelajaran merupakan salah satu komponen yang penting sebab akan menentukan sejauh mana tujuan pembelajaran telah tercapai. Tahap-tahap yang dilakukan dalam melaksanakan penilaian disajikan dalam tabel di bawah ini.

Tabel 1 Tahap-Tahap Pelaksanaan Penilaian

No	Tahap	Keterangan
1	Menentukan tujuan	Tujuan pembelajaran memiliki tekanan yang berbeda-beda sehingga penilaian yang dilakukan juga harus berbeda, misal dalam tes prestasi belajar. Lingkup materi yang ditanyakan atau diujikan disesuaikan dengan materi yang sudah di pelajari pada saat proses pembelajaran
2	Menentukan rencana penilaian	rencana penilaian berbentuk kisi-kisi matriks yang menggambarkan antara <i>behavioral objectives</i> (kemampuan yang menjadi sasaran dalam pembelajaran) dan <i>course content</i> (materi sajian yang dipelajari untuk mencapai kompetensi). Selain itu, rencana penilaian juga harus memperhatikan teknik penilaian yang akan digunakan dalam menilai keberhasilan peserta didik menguasai pembelajaran
3	Menyusun instrumen penilaian	Penyusunan instrumen penilaian disesuaikan dengan ranah yang akan diukur, teknik penilaian yang digunakan dan mengacu pada syarat instrumen yang baik. Syarat pokok instrumen yang baik yakni bersifat valid (sah) dan reliabel (dapat dipercaya)
4	Pengumpulan data dan informasi	Sebelum digunakan, instrumen penilaian diuji validitas dan reliabilitasnya. Kemudian, dapat

		dilakukan pengumpulan data yang dapat dilaksanakan pada setiap akhir pembelajaran secara terbuka dan objektif
5	Analisis dan interpretasi	Kegiatan ini hendaknya dilakukan setelah pengumpulan informasi. Diawali dengan tahap <i>skoring</i> sebagai tahap penentuan capaian penguasaan kompetensi oleh peserta didik. Interpretasi diwujudkan dalam bentuk deskripsi hasil penilaian tentang hasil belajar peserta didik
6	Tindak lanjut	Tindak lanjut dari hasil analisis dan interpretasi berupa rencana yang akan dilakukan pada pembelajaran selanjutnya untuk meningkatkan mutu pembelajaran.

Sumber : (Allivna, 2018: 27)

B. Pengembangan Instrumen Penilaian EHOTS

Pendidikan Sejarah bentuk pilihan ganda pada mata kuliah Sejarah Indonesia Masa Kolonial. Kualitas instrumen penilaian ditentukan oleh dua kriteria, yakni valid dan reliabel. Dalam mencapai kriteria tersebut, instrumen penilaian EHOTS harus melewati dua tahap penilaian. Penilaian tahap pertama dilakukan untuk menilai kevalidan instrumen tes yang dilakukan oleh dua orang ahli yakni ahli instrumen dan ahli evaluasi. Sementara itu, penilaian tahap kedua dilakukan uji coba terbatas yang melibatkan 20 mahasiswa program studi Pendidikan Sejarah.

Tahap tersebut difokuskan pada uji validitas, reliabilitas, tingkat kesukaran dan daya pembeda. Setiap tahap penilaian diakhiri dengan perbaikan berdasarkan temuan di lapangan. Proses yang dilakukan dalam pengembangan ini meliputi penyusunan produk soal tes EHOTS. Soal tes EHOTS yang telah dirancang selanjutnya dinilai oleh validator ahli instrumen, ahli evaluasi dan Dosen sejarah, kemudian tahap revisi untuk memperoleh produk awal soal tes EHOTS yang siap digunakan sebagai bahan uji coba terbatas. Hasil dari uji coba terbatas, sebagai bahan revisi untuk menjadi produk utama soal tes EHOTS yang siap digunakan sebagai bahan uji coba lapangan. Setelah diperoleh estimasi koefisien validitas, reliabilitas, tingkat kesukaran, dan daya pembeda dari hasil uji coba lapangan, maka akan diperoleh produk akhir soal tes EHOTS yang siap digunakan.

1) Data Hasil Validasi Ahli Instrumen

Validasi oleh ahli instrumen dilakukan sebagai salah satu syarat untuk memperoleh hasil penelitian (data) yang valid. Instrumen penelitian merupakan alat ukur pada penelitian, sehingga apabila instrumen dinyatakan valid, maka alat ukur yang digunakan dalam penelitian tersebut sudah valid. Validasi ini bertujuan untuk memperoleh masukan, saran perbaikan, dan sekaligus penilaian terhadap alat ukur yang digunakan dalam penelitian ini. Kegiatan validasi dilakukan dengan cara memberi naskah berupa lembar validasi kepada ahli instrumen. Ahli Instrumen dalam hal ini yakni bapak Dr. Samakmur, M.M. Selanjutnya dilakukan analisis penilaian instrumen penelitian sesuai dengan penilaian validator dengan menggunakan formula Aiken's V untuk menghitung *content validity*

coefficient. Berdasarkan hasil analisis menggunakan formula Aiken's V instrumen penelitian yang terdiri dari 10 butir daftar pernyataan, semuanya dinyatakan layak untuk digunakan. Walaupun demikian, beberapa pernyataan telah diperbaiki sesuai dengan saran dan masukan dari validator ahli instrumen penelitian.

2) Data Hasil Validasi Ahli Evaluasi

Validasi oleh ahli evaluasi dilakukan untuk melihat isi produk awal. Validasi ini bertujuan untuk mendapatkan masukan, saran perbaikan, dan sekaligus penilaian terhadap produk awal sebelum dilakukan uji coba terbatas. Kegiatan validasi dilakukan dengan cara memberi naskah produk awal yaitu berupa kisi-kisi instrumen soal dan soal tes EHOTS serta lembar validasi kepada ahli evaluasi. Ahli evaluasi dalam hal ini yakni ibu Dr. Hennilawati S.S, S.Pd., M.Pd. Selanjutnya dilakukan analisis penilaian butir soal tes EHOTS sesuai dengan penilaian validator dengan menggunakan formula Aiken's V untuk menghitung *content validity coefficient*. Berdasarkan hasil analisis menggunakan formula Aiken's V soal tes EHOTS yang terdiri dari 25 butir soal pilihan ganda semuanya dinyatakan layak untuk digunakan. Walaupun demikian, beberapa soal telah diperbaiki sesuai dengan saran dan masukan dari validator ahli evaluasi.

3) Data Hasil Validasi oleh Dosen Sejarah

Validasi oleh Dosen sejarah dilakukan untuk melihat isi dan keefektifan produk awal. Validasi ini bertujuan untuk mendapatkan masukan, saran perbaikan, dan sekaligus penilaian terhadap produk awal sebelum dilakukan uji coba terbatas. Kegiatan validasi dilakukan dengan cara memberi naskah produk awal yaitu berupa kisi-kisi instrumen soal dan soal tes EHOTS serta lembar validasi kepada Dosen sejarah. Selanjutnya dilakukan analisis penilaian butir soal tes EHOTS sesuai dengan penilaian validator dengan menggunakan formula Aiken's V untuk menghitung *content validity coefficient*. Hasil validasi dari Dosen Sejarah, diperoleh hasil bahwa semua butir soal pilihan ganda yang terdiri dari 25 butir soal berbasis EHOTS berada pada kategori valid atau layak digunakan dengan indeks terendah 0,58 dan tertinggi 0,91. Interpretasi ini dilakukan dengan menggunakan kriteria kurang dari 0,4 maka dikatakan validitasnya rendah, diantara 0,4-0,8 dikatakan validitasnya sedang (*mediocre*) dan jika lebih dari 0,8 dikatakan tinggi. Walaupun demikian, beberapa soal telah diperbaiki sesuai dengan saran dan masukan dari validator yakni Dosen sejarah.

4) Hasil Interpretasi Analisis Validitas Soal EHOTS

Validitas tes yang dilakukan dalam penelitian ini adalah uji validitas butir soal yang dikembangkan oleh peneliti. Butir soal terdiri dari 25 soal pilihan ganda berbasis EHOTS pada materi menganalisis perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda pada mata kuliah sejarah Indonesia

masa Kolonial. Hasil perhitungan kemudian diinterpretasikan dengan *r* tabel pada taraf signifikansi 5% dan $N=30$, maka diperoleh *r* tabel sebesar 0,361. Jika *r* hitung > *r* tabel maka butir soal dikatakan valid, sebaliknya jika *r* hitung kurang dari atau sama dengan *r* tabel maka butir soal dikatakan tidak valid. Berdasarkan hasil analisis pada soal EHOTS berbentuk pilihan ganda yang dikembangkan oleh peneliti terdapat 23 soal yang dinyatakan valid atau 80% dari keseluruhan soal.

Secara umum uji validitas soal EHOTS pada materi menganalisis perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda yang dilakukan pada mahasiswa semester III (tiga) dapat dilihat pada tabel di bawah ini.

Tabel : Uji Validitas Produk Awal Soal HOTS

No	Indeks Validitas	Butir Soal	Jumlah	Persentase
1	> 0,205 (valid)	1,2,3,6,7,8,9,10 11,12,13,14,15,17 18,19,20,21,22,23,24,25	23	80%
2	≤ 0,205 (Tidak valid)	4,5,16	3	20%

Sumber: data primer diolah dengan bantuan Microsoft Excel.

Berdasarkan tabel dan gambar di atas dapat disimpulkan jika soal EHOTS pada materi menganalisis perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda memiliki kualitas yang baik dari segi validitasnya karena jumlah butir soal yang valid lebih dari 50% dari keseluruhan soal. Artinya, soal HOTS tersebut merupakan soal yang dapat mengukur apa yang seharusnya diukur. Hal ini sejalan dengan pengertian validitas butir soal menurut Sudijono (2012: 163) bahwa validitas item dari suatu tes atau validitas butir soal adalah ketepatan mengukur yang dimiliki oleh sebuah item dalam mengukur apa yang seharusnya diukur.

5) Hasil Interpretasi Analisis Reliabilitas Soal EHOTS

Pengujian reliabilitas soal EHOTS pada materi menganalisis perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda, dilakukan secara manual dengan menggunakan rumus KR-20 dan bantuan program *microsoft Excel*. Soal EHOTS yang dikembangkan peneliti memiliki reliabilitas sebesar 0,58, sehingga dapat disimpulkan butir soal tersebut memiliki tingkat reliabilitas pada kategori cukup. Suatu tes dikatakan reliabel jika selalu memberikan hasil yang sama bila dites dengan kelompok yang sama pada waktu dan subjek yang berbeda (Arifin, 1991: 258). Hal ini sejalan dengan pernyataan Sudjana (2013: 16) bahwa reliabilitas alat penilaian menunjukkan keajegan alat tersebut dalam menilai apa yang hendak dinilainya. Dari penjelasan tersebut dapat disimpulkan bahwa soal EHOTS pada materi menganalisis perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari

ancaman Sekutu dan Belanda merupakan soal yang berkualitas cukup dari segi reliabilitasnya.

6) Hasil Interpretasi analisis Tingkat Kesukaran Soal EHOTS

Tingkat kesukaran butir soal dilihat dari besarnya indeks kesukaran. Hasil perhitungan indeks kesukaran diinterpretasikan dalam tiga kriteria, yaitu: $P = 0,00-0,30$ adalah soal yang sukar, $P = 0,31-0,70$ adalah soal dengan tingkat kesukaran sedang, dan $p = 0,71-1,00$ adalah soal yang mudah. Berdasarkan hasil perhitungan indeks kesukaran butir soal EHOTS dengan bantuan program *Microsoft Excel*, butir soal yang tergolong sukar berjumlah 3 dari 25 soal atau sebesar 12 %. Sementara butir soal yang memiliki tingkat kesukaran sedang berjumlah 17 atau sebesar 68%, dan butir soal yang tergolong mudah berjumlah 5 atau 20%. Secara umum tingkat kesukaran butir soal pilihan ganda dapat dilihat pada tabel di bawah berikut.

Tabel: Tingkat Kesukaran Produk Awal Soal Tes HOTS

No	Indeks Kesukaran	Butir Soal	Jumlah	Persentase
1	0,00-0,30 sukar	9,11,12	3	12%
2	0,31-0,70 sedang	2,3,4,5,7,8,10,13,14,15,16,18,20,21,22,23,24,25	17	68%
3	0,71-1,00 mudah	1,6,7, 17,19	5	20%

Sumber: Data Primer Diolah dengan menggunakan *Microsoft Excel*.

Suatu soal dianggap baik jika memiliki tingkat kesukaran sedang, yaitu antara 0,31-0,70 (Arikunto, 1999: 225). Dari gambar dan tabel di atas dapat disimpulkan bahwa soal EHOTS pada materi menganalisis perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda merupakan soal dengan kualitas yang cukup baik dilihat dari tingkat kesukarannya. Hal ini terlihat dari banyaknya soal yang memiliki tingkat kesukaran sedang yaitu 17 butir atau mencapai 68% dari keseluruhan soal.

7) Hasil Interpretasi Daya Pembeda Butir Soal EHOTS

Perhitungan Daya Pembeda dilakukan secara manual menggunakan program *Microsoft Excel* dengan membagi subjek menjadi dua bagian, 50% untuk kelompok atas dan 50% untuk kelompok bawah. Hasil perhitungan tersebut diinterpretasikan dalam lima kriteria, yaitu: (1) D = negatif, berarti soal tidak memiliki Daya Pembeda dan sebaiknya dibuang; (2) D = 0,00-0,19 berarti Daya Pembeda soal lemah; (3) D = 0,20-0,39 berarti Daya Pembeda soal cukup; (4) D = 0,40-0,69 berarti Daya Pembeda soal baik; dan D = 0,70-1,00 berarti Daya Pembeda soal baik sekali. Dari hasil perhitungan yang telah dilakukan, butir soal yang memiliki Daya Pembeda jelek berjumlah 5 atau sebesar 20%, butir soal yang memiliki Daya Pembeda cukup berjumlah 7 atau sebesar 28%, butir soal dengan Daya Pembeda baik

berjumlah 11 atau sebesar 44 %, dan butir soal yang memiliki Daya Pembeda kategori baik sekali berjumlah 2 atau 8%. Secara umum Daya Pembeda butir soal EHOTS bentuk pilihan ganda dapat dilihat pada tabel berikut.

Tabel : Daya Pembeda Produk Awal Soal EHOTS

No	Daya Pembeda	Nomor Butir Soal	Jumlah	Persentase
1	0,00-0,19 (jelek)	4,5,12,13,15	5	20%
2	0,20-0,39 (cukup)	2,7,8,16,18,21,22	7	28%
3	0,40-0,69 (baik)	1,3,6,9,10,14,17,19,20,23,25	11	44%
4	0,70-1,00 (baik sekali)	11,24	2	8%
5	Negatif (tidak memiliki daya pembeda)	-	0	0

Sumber: Data Primer diolah dengan *Microsoft Excel*

Arifin (1991: 273) menyatakan bahwa perhitungan daya pembeda adalah pengukuran sejauh mana suatu butir soal mampu membedakan peserta didik yang sudah menguasai materi dengan peserta didik yang belum atau kurang menguasai materi berdasarkan kriteria tertentu. Jadi dapat disimpulkan bahwa soal HOTS pada materi menganalisis perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda telah memiliki kualitas yang baik dilihat dari segi daya pembeda yaitu lebih dari 50% keseluruhan soal, sehingga dapat membedakan peserta didik kelompok atas dan kelompok bawah.

4. KESIMPULAN

Setelah melakukan semua tahapan penelitian maka dapat disimpulkan :

1. Produk akhir penelitian ini menghasilkan instrumen penilaian EHOTS pada materi perjuangan bangsa Indonesia dalam upaya mempertahankan kemerdekaan dari ancaman Sekutu dan Belanda. Instrumen penilaian EHOTS tersebut berupa perangkat soal tes HOTS yang terdiri dari 25 butir soal pilihan ganda dengan lima pilihan jawaban. Adapun teknik pengembangan instrumen penilaian dikembangkan melalui lima langkah yaitu (1) penelitian dan pengumpulan informasi, (2) perencanaan, (3) pengembangan produk awal, (4) uji coba terbatas, (5) revisi produk awal, yang diadopsi dari langkah pengembangan Borg & Gall.
2. Validitas soal tes HOTS (*Higher Order Thinking Skill*) ditunjukkan dari hasil analisis validator yaitu validator penilaian, validator instrumen dan Dosen Sejarah. Hasil analisis ahli penilaian menunjukkan bahwa instrumen penilaian EHOTS (*Higher Order Thinking Skill*) layak untuk digunakan. Sementara itu analisis validasi dari Dosen Sejarah juga menunjukkan bahwa instrumen penilaian HOTS (*Higher Order Thinking Skill*) yang terdiri dari 25 soal pilihan ganda layak untuk digunakan.

3. Karakteristik butir soal menunjukkan kualitas butir soal tes EHOTS. Kualitas tersebut diperoleh dari hasil analisis butir soal. Hasil perhitungan dari uji validitas soal EHOTS menunjukkan dari 25 item soal terdapat 3 soal yang tidak valid. Sementara tingkat reliabilitas soal HOTS adalah 0,5899 dengan kategori sedang atau cukup, tingkat kesukaran soal HOTS memiliki rata-rata 0,6168 dengan kategori cukup. Selanjutnya rata-rata daya pembeda yakni 0,2728 dengan kategori cukup.

5. REFERENSI

- Allivna. (2018). Pengembangan instrumen penilaian keterampilan proses dan aspek kognitif fisika siswa SMA berbasis local wisdom melalui proyek. *Tesis*, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
- Aman. (2011). *Model evaluasi pembelajaran sejarah*. Yogyakarta: Ombak
- Amirono & Daryanto. (2016). *Evaluasi dan penilaian pembelajaran kurikulum 2013*. Yogyakarta: Gava Media.
- Anderson, L.W & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy*. New York: Longman Publishing.
- Annisa, A.I. (2018). Pengembangan instrumen penilaian keterampilan kolaborasi an hasil belajar kognitif fisika peserta didik SMA melalui kerja lapangan berbasis kearifan lokal. *Tesis*, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
- Arifin, Z. (2009). *Evaluasi pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Arikunto, S (2002). *Dasar-dasar evaluasi pendidikan*. Jakarta: Bumi Aksara.
- Borg & Gall. (2003). *Education research*. New York: Allyn and Bacon.
- Brookhart, S. M. (2010). *How to assess high-er order thinking skills in your class- room*. Alexandria: ASCD
- Chiapetta, E.R & Koballa, T.L. (2010). *Science instruction in the middle and secondary schools: develoving fundamental knowledge and skills. (7th ed)*. Canada: Pearson.
- Chinedu, C.C., & Kamin, Y. (2015). Strategies for improving higher order thinking skill in teaching and learning of design and tecnology education. *Journal of Technical Education and Training (JTET)*, 35-43. ISSN 2229-8932.
- Dien, M., & Johan W. (2014). *Ilmu sejarah: sebuah pengantar*. Jakarta : Prenada Media Group
- Djaali & Muljono, P. (2008). *Pengukuran dalam bidang pendidikan*. Jakarta: PT. Gramedia Widiasarana Indonesia.
- Eko, P. W. (2014). *Penilaian hasil pembelajaran di sekolah*. Yogyakarta: Pustaka Pelajar.
- Febrian, R. (15 Agustus 2018). Soal-soal HOTS itu bikin pusing. *Tirto.id*. Diakses pada 15 Januari 2019, dari <https://tirto.id/soal-soal-hots-yang-bikin-siswa-pusing-itu-penting-cStV>.
- Ofianto. (2014). Model penilaian kemampuan berpikir historis (historical thinking) pendidikan sejarah sma dengan model rasch. *Dosertasi*, tidak diterbitkan. Universitas Negeri Yogyakarta.
- Osabede, U. P. (2014). Teachers assessment of classroom learning outcomes. *International Knowledge Sharing Platform*, 5(15), 15-22.
- Paul, R., & Nosich, G. M. (1993). A model for the national assessment of higher order thinking. Foundation for critical thinking. Diakses pada tanggal 20 Januari 2019 pukul 08.00 WIB dari <https://www.criticalthinking.org/pages/a-model-for-the-national-assessment-of-higher-order-thinking/591>.
- Pratiwi, U. (2015). Pengembangan insrumen penilaian HOTS berbasis kurikulum 2013 terhadap sikap disiplin. *Jurnal penelitian dan pembelajaran IPA*, 1 (1),