



PENGEMBANGAN INSTRUMEN PENILAIAN OTENTIK PEMBELAJARAN FISIKA UNTUK MENGUKUR KETERAMPILAN PROSES SAINS DAN BERPIKIR KRITIS

Amalia Khasanah*, Universitas Negeri Yogyakarta, Indonesia
Zuhdan Kun Prasetyo, Universitas Negeri Yogyakarta, Indonesia

*e-mail: lia.amaliakhasanah96@gmail.com

Abstrak. Penelitian ini bertujuan untuk (1) menghasilkan instrumen penilaian otentik yang layak digunakan untuk mengukur keterampilan proses sains dan keterampilan berpikir kritis peserta didik, (2) mengetahui seberapa besar ketercapaian keterampilan proses sains dan berpikir kritis peserta didik ditinjau dari hasil belajar aspek kognitif. Desain Penelitian yang digunakan adalah penelitian pengembangan modifikasi model Wilson dan Oriondo dan Antonio. Analisis validitas isi instrumen menggunakan CVR/CVI, validitas empiris dan data politomus hasil uji coba dianalisis menggunakan model *Rasch* dengan bantuan program Winsteps 3.37. Hasil penelitian menunjukkan bahwa (1) telah dihasilkan instrumen penilaian otentik yang layak untuk mengukur keterampilan proses sains dan berpikir kritis peserta didik. Validitas isi instrumen oleh ahli dan praktisi termasuk dalam kategori sangat baik. Validitas empiris berdasarkan tiga kriteria yaitu OUTFIT MNSQ, ZSTD, Pt Mean Corr untuk 8 butir instrumen keterampilan proses sains dan 5 butir instrumen keterampilan berpikir kritis, 1 butir instrumen keterampilan proses sains tidak memenuhi kriteria. Reliabilitas instrumen berdasarkan *index sparation item* sangat bagus dengan nilai $>0,90$, (2) ketercapaian keterampilan proses sains peserta didik sebagian besar berada pada kategori sangat tinggi 31,25% dan rendah 21,88%. Ketercapaian peserta didik untuk keterampilan berpikir kritis sebagian besar berada pada kategori tinggi 28,12%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa instrumen penilaian otentik yang dikembangkan dapat digunakan sebagai salah satu instrumen alternatif untuk mengukur keterampilan proses sains dan berpikir kritis peserta didik.

Kata Kunci: *penilaian otentik, keterampilan proses sains, keterampilan berpikir kritis, dan hasil belajar.*

Abstract. The research aims to 1) generate feasible authentic assessment instrument to measure learners' science process and critical thinking skills, 2) know how far the learners' science process and critical thinking skills achievement in terms of cognitive aspect. The used research design was the modification of Wilson and Oriondo and Antonio model. Instrument validity was analyzed by using CVR/CVI, empirical validity and the polytomy data was analyzed using Winsteps 3.37. Results of the research indicate that (1) a feasible authentic assessment instrument to measure learners' science process and critical thinking skills is already generated. The authentic assessment instrument in terms of content validity is categorized as very, meanwhile for the empirical validation is based on the three criteria that are OUTFIT MNSQ, ZSTD, Pt Mean Corr for 8 points of science process skill and 5 points of critical thinking skill instrument, yet 1 point of instrument of science process skill does not meet the criteria. The instrument reliability of science process and critical thinking based on index sparation item is very good with >0.90 , and (2) the learner's achievement of the science process skill is mostly on the categorize of very high with 31.25% and low with

21.88%. The learner's achievement regarding the critical thinking skill is mostly on the categorize of 28.12%. Based on the result, authentic assesment instrument that has been developed can be use as one of the alternative assesment instrument to measure learners' science process and critical thinking skills.

Keywords: *authentic sssessment, science process skill, critical thinking skill, and learning outcome.*

PENDAHULUAN

Penilaian atau evaluasi merupakan salah satu hal terpenting yang harus dilakukan dalam proses pembelajaran, sebagaimana yang dijelaskan oleh Temiz *et al*, (2006) menyatakan bahwa penilaian dan pembelajaran merupakan dua sisi mata uang yang tidak dapat dipisahkan. Hal tersebut sesuai dengan Peraturan Menteri Nomor 23 Tahun 2016 Tentang Standar Penilaian Pendidikan yang menyebutkan bahwa salah satu prinsip penilaian yaitu terpadu, yang berarti bahwa penilaian merupakan salah satu komponen yang tak terpisahkan dari kegiatan pembelajaran. Penilaian otentik (*authentic assesment*) memiliki relevansi kuat terhadap pendekatan ilmiah dalam pembelajaran sesuai dengan tuntutan Kurikulum 2013. Peraturan Menteri Nomor 22 tahun 2016 menghendaki bahwa penilaian proses pembelajaran menggunakan pendekatan penilaian otentik (*authentic assesment*) yang menilai kesiapanpeserta didik, proses, dan hasil belajar secara utuh.

Penilaian otentik (*authentic assesment*) adalah pengukuran yang bermakna atas hasil belajar peserta didik untuk ranah sikap, keterampilan, dan pengetahuan. Istilah otentik merupakan sinonim dari asli, nyata, atau valid. Penilaian otentik sering dikontradiksikan dengan penilaian yang menggunakan standar tes berbasis norma, pilihan ganda, benar-salah, menjodohkan, atau membuat jawaban singkat karena bentuk-bentuk penilaian tersebut dirasa kurang mampu memberikan gambaran secara nyata mengenai proses berpikir peserta didik dan sejauh mana mereka memahami dan menguasai materi yang disampaikan dalam pembelajaran.

Evaluasi yang berbasis *authentic assesment* mempunyai tujuan agar peserta didik mampu berfikir melalui sains dan membentuk kemampuan berpikir tingkat tinggi. Pada kurikulum 2013 keterampilan proses sains dan kemampuan berpikir tingkat tinggi merupakan dua hal yang tidak kalah penting. Keterampilan proses sains merupakan salah satu aspek dari sains. Keterampilan proses sains digunakan untuk membangun tubuh pengetahuan yang merupakan kumpulan dari ilmu pengetahuan. Selain keterampilan proses sains, kemampuan berpikir tingkat tinggi yaitu berpikir kritis perlu dimiliki peserta didik karena dapat membantu peserta didik untuk mengembangkan potensi intelektualnya, sebagaimana dijelaskan oleh Johnson, (2007) menyatakan bahwa kemampuan berpikir kritis sangat diperlukan bagi seseorang, hal ini dianggap penting karena mampu membantu peserta didik mengembangkan potensi intelektualnya, memiliki kemampuan untuk mengevaluasi secara sistematis, dan mampu berpendapat secara terorganisasi. Menurut Lambertus (2009) definisi berpikir kritis paling sedikit memuat tiga hal. Pertama, berpikir kritis merupakan proses pemecahan masalah dalam suatu konteks interaksi dengan diri sendiri, dunia orang lain dan atau lingkungannya. Kedua, berpikir kritis merupakan proses penalaran reflektif berdasarkan informasi dan kesimpulan yang telah diterima sebelumnya yang hasilnya terwujud dalam penarikan kesimpulan. Ketiga, berpikir kritis berakhir pada keputusan apa yang diyakini dan dikerjakan. Apriza dan Mahmudi (2015) menyatakan bahwa berpikir kritis sangat penting untuk mempelajari materi baru guna mengkaitkannya dengan apa yang telah peserta didik ketahui.

Pada implementasinya, khususnya pembelajaran fisika masih sering dijumpai soal-soal evaluasi pada penilaian formatif yang cenderung menuntut kemampuan ingatan dan

memecahkan masalah lewat solusi matematis dalam bentuk soal pilihan ganda yang sering dikontradiksikan dengan penilaian otentik, sehingga cenderung tidak efektif apabila digunakan sebagai alat evaluasi untuk mengukur kemampuan berpikir tingkat tinggi pada peserta didik. Wiggins (1990) menyatakan bahwa suatu penilaian dapat dikatakan otentik ketika pendidik menguji peserta didik secara langsung dengan tugas intelektual yang terpercaya. Bentuk penilaian melibatkan peserta didik pada persoalan yang berguna atau pertanyaan penting sehingga peserta didik harus menggunakan pengetahuan untuk menunjukkan kinerja secara efektif dan kreatif. Oleh karena itu, dalam penelitian ini, peneliti bermaksud untuk melakukan penelitian pengembangan instrumen penilaian otentik yang berbentuk tes tertulis berupa soal uraian yang dapat digunakan untuk mengukur keterampilan proses sains pada tingkatan HOTS (*Higher Order Thinking Skill*) dan keterampilan berpikir kritis. Tes tertulis berbentuk essay/uraian dipilih sebagaimana yang dijelaskan oleh Nitko & Brookhart, (2011) menyatakan tes essay lebih memberikan kesempatan kepada peserta didik untuk menampilkan kemampuan dalam menulis, mengorganisasikan, mengekspresikan, dan menjelaskan hubungan antar ide sehingga mampu menilai kemampuan berpikir tingkat tinggi peserta didik.

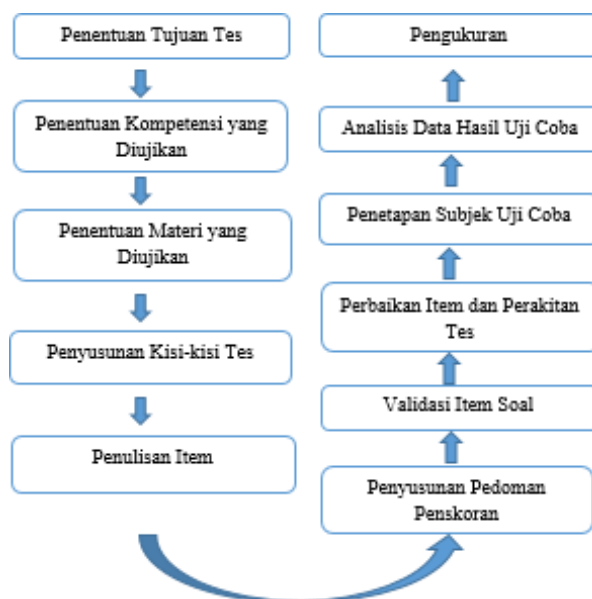
Peneliti telah melakukan observasi pada salah satu sekolah menengah atas di kabupaten Sleman yaitu di MAN 3 Sleman. Berdasarkan observasi, didapatkan informasi bahwa belum pernah dilakukan penelitian terkait dengan pengembangan instrumen penilaian otentik pembelajaran fisika untuk mengukur keterampilan proses sains dan berpikir kritis peserta didik di MAN 3 Sleman sebelumnya. Oleh karena itu, adanya penelitian untuk mendukung perlu dilakukan. Penelitian ini akan menjadi sangat penting artinya bagi kegiatan pembelajaran fisika di MAN 3 Sleman agar keterampilan proses sains dan berpikir kritis peserta didik dapat terukur.

Masalah utama yang akan diungkap dalam penelitian ini adalah apakah instrumen penilaian otentik yang dikembangkan layak untuk mengukur keterampilan proses sains dan berpikir kritis peserta didik? dan berapakah ketercapaian keterampilan proses sains dan keterampilan peserta didik ditinjau dari hasil belajar aspek kognitif?

Sejalan dengan rumusan masalah yang akan diselesaikan, maka tujuan penelitian ini adalah: (1) menghasilkan instrumen penilaian otentik yang layak digunakan untuk mengukur keterampilan proses sains dan keterampilan berpikir kritis peserta didik, dan (2) mengetahui seberapa besar ketercapaian keterampilan proses sains dan berpikir kritis peserta didik ditinjau dari hasil belajar aspek kognitif.

METODE

Model pengembangan instrumen tes keterampilan proses sains dan berpikir kritis menggunakan modifikasi model Wilson dan Oriundo dan Antonio. Tahapan pengembangan instrumen tes terdiri atas tiga tahap, yakni (1) tahap perancangan tes yang terdiri dari: penentuan tujuan tes, penentuan kompetensi yang diujikan, penentuan materi yang diujikan, penyusunan kisi-kisi tes, penulisan item, validasi item, perbaikan item dan perakitan tes, serta penyusunan pedoman penskoran; (2) ujicoba tes yang terdiri dari: penentuan subjek uji coba, pelaksanaan uji coba, dan analisis data hasil uji coba; dan (3) perakitan tes (Istiyono, Mardapi, & Suparno, 2014: 6).



Gambar 1. Skema prosedur penelitian Penelitian dilakukan di MAN 3 Sleman pada bulan April sampai dengan Mei 2018.

Perancangan tes yang berupa penentuan tujuan tes, penentuan kompetensi yang diujikan, penentuan materi yang diujikan, penyusunan kisi-kisi tes, penulisan item, validasi item, perbaikan item dan perakitan tes, serta penyusunan pedoman penskoran dilakukan pada bulan Oktober sampai dengan Maret 2018.

Subjek penelitian meliputi peserta didik MAN 3 Sleman. Uji keterbacaan soal melibatkan peserta didik kelas XI MAN 3 Sleman dan uji coba instrumen tes melibatkan 32 peserta didik kelas X MAN 3 Sleman.

Instrumen penelitian berupa perangkat penunjang pembelajaran yang berupa RPP, dan pengumpul data yang berupa instrumen penilaian otentik keterampilan proses sains, instrumen penilaian otentik keterampilan berpikir kritis, lembar validasi instrumen, dan lembar keterlaksanaan RPP.

Analisis data menggunakan analisis kualitatif dan kuantitatif. Analisis kualitatif bertujuan untuk merevisi produk instrumen penilaian berdasarkan saran validator dan data uji coba. Analisis kuantitatif digunakan untuk mengetahui validitas dan reliabilitas instrumen penilaian yang dikembangkan.

Analisis validitas instrumen tes menggunakan dihitung menggunakan *content validity ratio* (CVR) dan *content validity index* (CVI). Pemberian skor pada item divalidasi dengan CVR. Cara menghitung nilai *content validity ratio* (CVR) adalah dengan persamaan:

$$CVR = \frac{(N_e - \frac{N}{2})}{\frac{N}{2}}$$

dengan:

N_e = jumlah validator yang menyetujui

N = jumlah total validator

Kemudian dihitung pula *content validity index* (CVI) yang merupakan indikasi isi tes. CVI merupakan rata-rata nilai CVR dari semua item.

$$CVI = \frac{\text{jumlah CVR}}{\text{jumlah butir validasi instrumen}}$$

Rentang hasil nilai CVR dan CVI adalah $1 < 0 < 1$.

Angka tersebut dikategorikan sebagai berikut:

$-1 < x < 0$ = tidak baik

0 = baik

$0 < x < 1$ = sangat baik

Analisis kelayakan RPP menggunakan skala interval 1-5 dan dianalisis menggunakan analisis Simpangan Baku Ideal (*SBi*), dengan cara menghitung nilai \bar{X} dimana:

$$\bar{X} = \frac{\sum x}{n}$$

\bar{X} : nilai rata-rata

$\sum x$: jumlah total jawaban nilai validator

n : jumlah validator

Kemudian menghitung rata-rata ideal dengan rumus:

$$\bar{X} = \frac{1}{2} (\text{skor maksimal ideal} + \text{skor minimal ideal})$$

Dan menghitung simpangan baku ideal dengan rumus:

$$SBi = \frac{1}{6} (\text{skor maksimal ideal} + \text{skor minimal ideal})$$

Kriteria penilaian ditentukan dengan kategorisebagai berikut:

Tabel 1. Kategori penilaian skala lima

Rentang Nilai	Kategori
$\bar{X} > \bar{X} + 1,8 SBi$	Sangat Baik
$\bar{X} + 0,6 SBi < \bar{X} \leq \bar{X} + 1,8 SBi$	Baik
$\bar{X} - 0,6 SBi < \bar{X} \leq \bar{X} + 0,6 SBi$	Cukup Baik
$\bar{X} - 1,8 SBi < \bar{X} \leq \bar{X} - 0,6 SBi$	Kurang Baik
$\bar{X} \leq \bar{X} - 1,8 SBi$	Sangat Kurang

Keterangan:

\bar{X} : skor aktual

\bar{X} : rerata skor ideal

SBi : simpangan baku ideal

Pada RPP juga dilakukan analisis *PA* (*Percentage Agreement*) untuk mengetahui kekonsistenan penilaian antar validator. Menurut Borich (1994: 385) nilai *PA* dapat diketahui dengan menggunakan persamaan berikut.

$$PA = \left(1 - \frac{A - B}{A + B}\right) \times 100\%$$

keterangan:

PA : *percentage of agreement*

A : skor validator yang lebih tinggi

B : skor validator yang lebih rendah

Instrumen dikatakan baik apabila *percentage of agreement* $\geq 75\%$.

Analisis keterlaksanaan pembelajaran digunakan untuk mengetahui apakah semua kegiatan dapat terlaksana dengan baik. Adapun hasil dari observer mengenai pengamatan pembelajaran tersebut dianalisis dengan menghitung *interjudge agreement* (IJA) dengan persamaan:

$$IJA = \frac{A_Y}{A_Y + A_N} \times 100\%$$

Keterangan:

A_Y : kegiatan yang terlaksana

A_N : kegiatan yang tidak terlaksana

Kriteria pembelajaran dikatakan layak apabila nilai IJA lebih dari 75%.

Data politomus hasil uji coba dianalisis menggunakan model *Rasch* yang merupakan pengembangan model analisis oleh Georg Rasch dari teori respon butir 1 PL (satu Parameter Logistic) dengan bantuan program *Winsteps 3.37*. Kesesuaian item (*item fit*) dengan model *rasch* dapat menjelaskan apakah item/butir instrumen berfungsi normal dalam melakukan pengukuran atau tidak. Kriteria yang digunakan untuk memeriksa kesesuaian butir instrumen untuk dapat dikatakan sesuai atau tidak dengan model yaitu dengan melihat nilai *Outfit Mean Square* (MNSQ), *Outfit Z-standard* (ZSTD), dan *Point Measure Correlation* (Pt Mean Corr).

Tabel 2. Kriteria Validitas Butir Instrumen

Kriteria	Koefisien	Keterangan
OUTFIT	$0,5 < \text{MNSQ} < 1,5$	Diterima
MNSQ		
OUTFIT	$-2,0 < \text{ZSTD} < +2,0$	Diterima
ZSTD		
Pt. Mean Corr	$0,4 < \text{Pt. Mean Corr} < 0,85$	Diterima
Mean Corr		
Corr		

Kriteria reliabilitas butir instrumen didasarkan pada indeks sparasi item dan pendapat Suryabrata (2002), bahwa hasil uji coba instrumen yang memiliki reliabilitas sekurang-kurangnya 0,90 dapat digunakan untuk mengambil keputusan tentang individu.

Analisis tingkat kesukaran butir menggunakan kriteria menurut Hambleton & Swaminathan (1985: 36) menyatakan bahwa item dikatakan baik jika tingkat kesukarannya lebih dari -2,0 atau kurang dari +2,0 ($-2,0 < \text{difficulty} < +2,0$).

Analisis tingkat kemampuan peserta didik ditunjukkan dengan nilai *logit* pada *person measure* dengan program Winsteps 3.37. Skor hasil yang diperoleh peserta didik dari instrumen keterampilan proses sains dan instrumen berpikir kritis berupa angka kemudian dikonversi menjadi skala lima dan dianalisis menggunakan Simpangan Baku Ideal (*Sbi*) untuk mengetahui tingkat ketercapaian peserta didik.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Hasil pengembangan instrumen meliputi rumusan kisi-kisi penilaian otentik, kisi-kisi keterampilan proses sains, dan kisi-kisi keterampilan berpikir kritis. Perumusan kisi-kisi tersebut digunakan sebagai acuan dalam penyusunan indikator instrumen tes dan lembar validasi instrumen. Rumusan kisi-kisi yang digunakan dalam pengembangan instrumen tes pada penelitian ini disajikan pada tabel-tabel berikut.

Tabel 3. Kisi-kisi Penilaian Otentik

Hasil Belajar	Tes tertulis	Ranah kognitif	Indikator
Hasil belajar kognitif	Uraian	Menganalisis (C4)	<ul style="list-style-type: none"> - Menganalisis data dalam pemecahan suatu masalah - Menganalisis argumen/pendapat tentang suatu istilah, fenomene/gejala - Mengaitkan ide berdasarkan pertimbangan observasi
		Mengevaluasi (C5)	<ul style="list-style-type: none"> - Memberi argumentasi tentang suatufenomena atau gejala - Menginterpretasi data yang disajikan dalam bentuk grafik, diagram dan/atau menggunakannya untuk menjawab pertanyaan
		Menciptakan (C6)	Menyusun/Membuat strategi/langkah yang rasional untuk memecahkan suatu masalah

Tabel 4. Kisi-kisi Ketrampilan Proses Sains

Aspek	Sub Aspek	Indikator
Menginvestigasi	Menganalisis data untuk menggambarkan kesimpulan dalam pemecahan suatu masalah	Melakukan analisis dalam pemecahan suatu masalah
Menginterpretasi	Membaca grafik, diagram dan/atau menggunakan lainnya untuk menjawab pertanyaan	Membaca informasi yang disajikan dalam bentuk grafik, diagram dan/atau menggunakannya untuk menjawab pertanyaan
Membuat model	Mengembangkan representasi secara fisik atau metal untuk menjelaskan suatu ide dalam perencanaan percobaan sederhana	<ul style="list-style-type: none"> - Menyusun langkah-langkah percobaan untuk membuktikan suatu prinsip atau hukum - Membuat skema alat percobaan untuk membuktikan suatu prinsip atau hukum

Tabel 5. Kisi-kisi Keterampilan Berpikir Kritis

Aspek	Sub Aspek	Indikator
Klarifikasi dasar	Memberikan penjelasan sederhana	<ul style="list-style-type: none"> - Menganalisis pendapat/ argumen tentang suatu istilah, fenomena atau Gejala - Memberi argumentasi tentang suatu fenomena atau gejala
Membangun keterampilan dasar	Melakukan pertimbangan observasi	Mengkaitkan ide berdasarkan pertimbangan observasi
Klarifikasi lanjut	Mendefinisikan istilah dan mempertimbangkan definisi	Membuat sebuah definisi dari suatu istilah berdasarkan pengamatan yang

Mengatur strategi dan taktik	menggunakan kriteria yang tepat	dilakukan
	Menentukan suatu tindakan untuk menemukan solusi alternatif dari suatu permasalahan	Menyusun strategi/langkah yang rasional untuk memecahkan suatu masalah

Setiap aspek dibuatkan minimal satu indikator butir instrumen. Satu indikator minimal satu butir instrumen. Instrumen penilaian otentik keterampilan proses sains terdiri dari 8 butir soal dan instrumen penilaian otentik keterampilan berpikir kritis terdiri dari 5 butir soal. Persebaran materi pada butir instrumen dapat dilihat pada Tabel 6 dan 7 berikut ini.

Tabel 6. Persebaran Materi pada Butir Instrumen Keterampilan Proses Sains

Aspek	Sub Aspek	Materi			
		Usaha	Energi	Konsep Usaha - energi	HKE
Menginvestigasi	Menganalisis	No 1, No 2	-	No 4	No 3
Menginterpretasi	Membaca grafik	-	No 5, No 6, No 7	-	-
Memuat model	Mengembangkan representasi	-	-	-	No 8

Tabel 7. Persebaran Materi pada Butir Instrumen keterampilan Berpikir Kritis

Aspek	Sub Aspek	Materi			
		Usaha	Energi	Konsep Usaha - energi	HKE
Klarifikasi Dasar	Memberikan penjelasan sederhana	No 1	No 2	-	-

Membangun Keterampilan Dasar	Melakukan pertimbangan observasi	-	-	No 3	-
Klarifikasi Lanjut	Men-definisikan istilah	No 4	-	-	-
Mengtur Strategi dan Taktik	Meneentukan tindakan	No 5	-	-	-

Hasil validasi isi masing-masing instrument penilaian otentik disajikan pada Tabel 8 berikut.

Tabel 8. Hasil validasi kelayakan instrumen penilaian otentik

Instrumen	CVI	Kategori
Keterampilan proses sains	0.99	Sangat baik
Keterampilan berpikir kritis	0.99	Sangat baik

Berdasarkan hasil analisis yang ditunjukkan oleh Tabel 8, instrumen tes yang dikembangkan memiliki CVI dengan kategori sangat baik sehingga dikatakan layak digunakan untuk pengambilan data penelitian.

Penilaian kelayakan perangkat penunjang pembelajaran berupa RPP dengan *SBi* diperoleh nilai rata-rata 4,88 dengan kategori sangat baik. Keterlaksanaan RPP pada masing-masing pertemuan disajikan dalam Tabel 9 berikut.

Tabel 9. Hasil analisis keterlaksanaan RPP

	Pertemuan 1	Pertemuan 2	Pertemuan 3
Nilai IJA	80,76%	87,28%	100%
Kategori	Sangat baik	Sangat baik	Sangat baik
Rata-rata		89,35%	
Kategori		Sangat baik	

*Analisis dengan IJA: *Interjudge Agreement*

Hasil tersebut didukung oleh nilai *Percentage of Agreement* yang diperoleh dari kedua observer yaitu sebesar 97,51% yang berarti bahwa kedua validator memiliki persepsi yang sama.

Berdasarkan analisis butir instrumen menggunakan program Winsteps 3.37, dapat diketahui kualitas butir instrumen penilaian. Analisis instrumen dengan *Rasch model* digunakan untuk mengetahui kesesuaian item tes (*goodness of fit*) dengan model *Rasch*. Parameter fit statistik dengan model *Rasch* disajikan pada Tabel 10 dan Tabel 11 berikut.

Tabel 10. Parameter Fit Statistik Instrumen Keterampilan Proses Sains

No	Parameter Tes	Estimasi
1	<i>OUTFIT MNSQ</i>	0,81 – 1,70
2	<i>ZSTD</i>	-0,7 – 2,3
3	<i>Pt Mean Corr</i>	0,28 – 0,70

Hasil analisis menunjukkan bahwa butir instrumen nomor 8 untuk parameter *OUTFIT MNSQ*, *ZSTD*, dan *Pt Mean Corr* ketiganya tidak terpenuhi karena nilainya tidak berada pada rentang kriteria yang telah ditetapkan, sehingga untuk perakitan tes selanjutnya butir instrumen ini perlu diperbaiki atau diubah.

Tabel 11. Parameter Fit Statistik Instrumen Keterampilan Berpikir Kritis

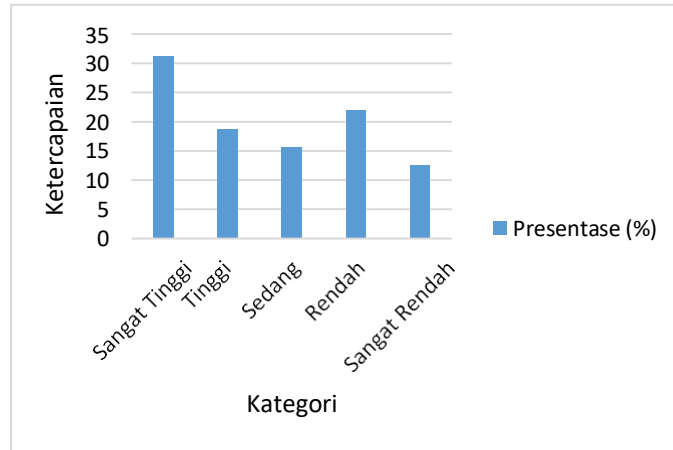
No	Parameter Tes	Estimasi
1	<i>OUTFIT MNSQ</i>	0,68 – 1,30
2	<i>ZSTD</i>	-1,2 – 0,9
3	<i>Pt Mean Corr</i>	0,01 – 0,69

Hasil analisis menunjukkan 7 butir instrumen penilaian otentik keterampilan proses sains dan 5 butir instrumen penilaian otentik keterampilan berpikir kritis fit terhadap *Rasch model*.

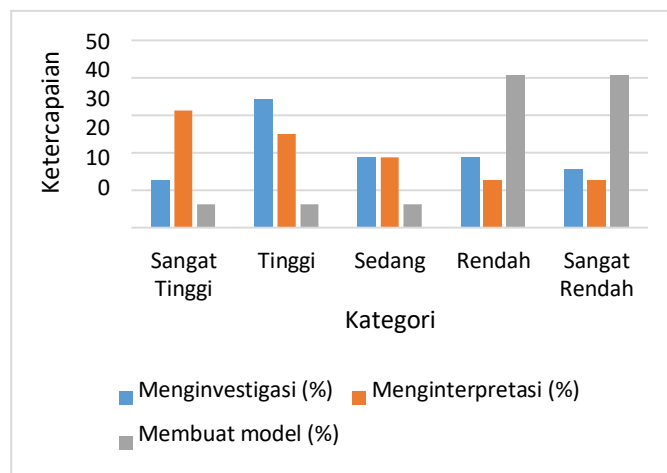
Reliabilitas instrumen keterampilan proses sains berdasarkan item (*indeks sparasi item*) yaitu 0,97 sedangkan reliabilitas instrumen keterampilan berpikir kritis yaitu 0,98. Reliabilitas kedua instrumen tersebut berada dalam kategori istimewa.

Tingkat kesukaran butir instrumen keterampilan proses sains terletak pada rentang - 0,63 sampai dengan 0,68, sedangkan untuk instrumen keterampilan berpikir kritis *difficulty index* butir instrumen terletak pada rentang - 0,81 sampai 0,32. Dengan demikian, berdasarkan *difficulty index* semua butir instrumen keterampilan proses sains dan instrumen keterampilan berpikir kritis tergolong baik karena memenuhi kriteria $-2,00 < \text{difficulty index} < +2,00$.

Hasil analisis tingkat ketercapaian keterampilan proses sains peserta didik secara keseluruhan dan tingkat ketercapaian tiap aspek disajikan pada Gambar 2 dan Gambar 3 berikut.

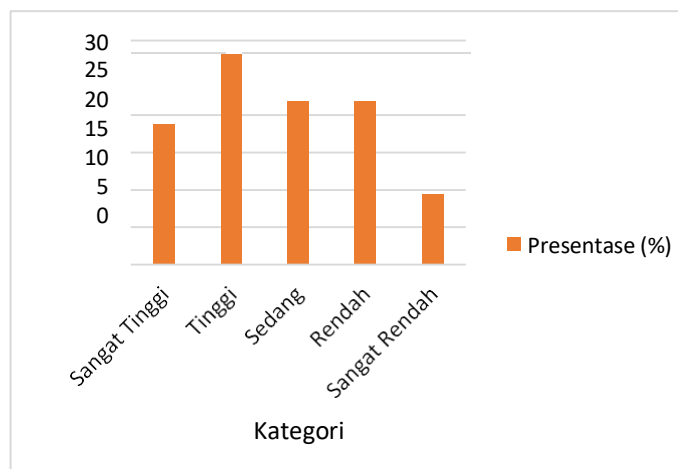


Gambar 2. Presentase Kategori Ketercapaian Keterampilan Proses Sains Peserta Didik

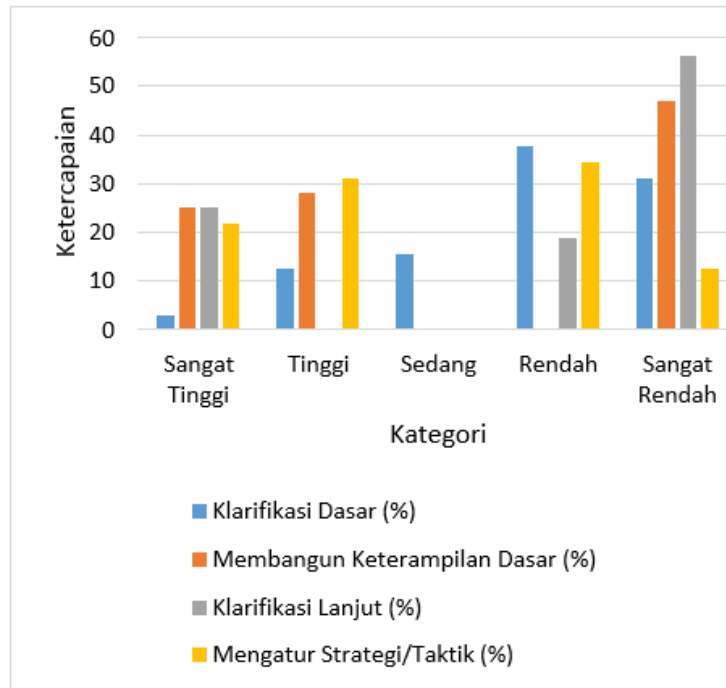


Gambar 3. Presentase Kategori Ketercapaian Keterampilan Proses Sains Peserta Didik Tiap Aspek

Hasil analisis tingkat ketercapaian keterampilan berpikir kritis peserta didik secara keseluruhan dan tingkat ketercapaian tiap aspek disajikan pada Gambar 4 dan Gambar 5 berikut



Gambar 4. Presentase Kategori Ketercapaian Keterampilan Berpikir Kritis Peserta Didik



Gambar 5. Presentase Kategori Ketercapaian Keterampilan Berpikir Kritis Peserta Didik Tiap Aspek

Pembahasan

1. Tingkat Kelayakan Instrumen Penilaian Otentik untuk Mengukur Keterampilan Proses Sains dan Berpikir Kritis

Hasil analisis menggunakan CVR/CVI masing-masing butir instrumen pada instrumen penilaian otentik keterampilan proses sains dan instrumen penilaian otentik keterampilan berpikir kritis yang dikembangkan dinyatakan valid. Berdasarkan hasil uji coba lapangan luas, 5 butir instrumen keterampilan berpikir kritis telah memenuhi kriteria *OUTFIT MNSQ*, *ZSTD*, dan *Pt Mean Corr*, yang berarti bahwa instrumen penilaian fit dengan model yaitu model *Rasch* 1 parameter logistik (1-PL), sedangkan pada instrumen keterampilan proses sains terdapat 1 item yang tidak fit dengan model yaitu butir instrumen nomor 8, sehingga butir ini tidak layak digunakan. Validitas instrumen yang baik ini disebabkan oleh beberapa faktor yaitu:

- Instrumen dikembangkan sesuai dengan prosedur pengembangan yang digunakan yaitu pengembangan modifikasi model Wilson dan model oriondo dan Antonio.
- Tiap butir instrumen dari instrumen penilaian otentik keterampilan proses sains dikembangkan berdasarkan kisi-kisi penilaian otentik dan kisi-kisi keterampilan proses sains. Sama halnya dengan instrumen penilaian otentik keterampilan proses sains, tiap butir instrumen penilaian otentik keterampilan berpikir kritis juga dikembangkan berdasarkan kisi-kisi penilaian otentik yang digabungkan dengan kisi-kisi keterampilan berpikir kritis. Item yang dikembangkan pada kedua instrumen disesuaikan dengan materi usaha dan energi yang sesuai dengan perkembangan peserta didik kelas X MIA SMA/MA.
- Peserta didik yang mengikuti uji coba dalam mengerjakan soal sungguh-sungguh karena nilai yang diperoleh dijadikan sebagai nilai ulangan harian, serta kecurangan peserta didik dalam mengerjakan soal diminimalisir dengan mengatur tempat duduk peserta didik berdasarkan urutan nomor presensi.

Reliabilitas instrumen keterampilan proses sains dan instrumen keterampilan berpikir

kritis berdasarkan *index sparation item* cukup tinggi yakni $>0,90$, hasil ini menunjukkan bahwa pengukuran menggunakan instrumen yang dikembangkan reliabel.

2. Tingkat Ketercapaian Keterampilan Proses Sains dan Berpikir Kritis Peserta Didik

Hasil analisis ketercapaian keterampilan proses sains dan keterampilan berpikir kritis peserta didik menggunakan rerata peserta didik dalam menjawab tiap butir instrumen. Ketercapaian keterampilan proses sains peserta didik secara keseluruhan dominan berada pada kategori sangat tinggi yaitu 31,25%. Hasil tersebut jika dianalisis lebih rinci dapat memberikan informasi bahwa ketercapaian untuk aspek menginvestigasi sebagian besar termasuk dalam kategori tinggi yaitu sebanyak 34,38%, ketercapaian peserta didik untuk aspek menginterpretasi sebagian besar termasuk dalam kategori sangat tinggi yaitu sebesar 31,25%, dan ketercapaian untuk aspek membuat model sebagian besar termasuk dalam kategori rendah dan sangat rendah yaitu sebesar 40,62%. Berdasarkan hasil analisis yang diperoleh, dapat diketahui bahwa instrumen penilaian otentik keterampilan proses sains yang dikembangkan mampu mengukur kemampuan peserta didik dari yang paling tinggi hingga kemampuan yang paling rendah. Aspek membuat model memiliki nilai ketercapaian yang paling rendah dibandingkan dengan dua aspek lainnya yaitu menginvestigasi dan menginterpretasi. Hal ini dikarenakan aspek membuat model berada pada tingkatan taksonomi Bloom paling atas yaitu mencipta (C6) yang juga merupakan tingkatan HOTS teratas. Sebagian kecil peserta didik, yaitu sebanyak 6,25% mampu menjawab soal kategori C6 ini dengan sangat baik, namun sebagian besar peserta didik yang lainnya belum mampu menjawab dengan baik. Dengan demikian, penilaian otentik yang dilakukan ini dapat memberikan gambaran nyata mengenai kemampuan peserta didik yang sesungguhnya. Hal ini sesuai dengan Azim dan Khan (2012) yang menyatakan bahwa tes dengan menggunakan konteks nyata yang membutuhkan keterampilan berpikir kompleks dan berbagai macam metode penyelesaian (HOTS) merupakan salah satu strategi dalam penilaian otentik. HOTS dapat dideskripsikan sebagai keterampilan berpikir yang tidak sekedar memerlukan kemampuan untuk mengingat, tetapi juga kemampuan berpikir tingkat tinggi seperti keterampilan menganalisis, mengevaluasi, dan mencipta (Yulianti, et al, 2018).

Ketercapaian keterampilan berpikir kritis peserta didik secara keseluruhan dominan pada kategori tinggi yaitu 28,12%. Berdasarkan hasil analisis ketercapaian keterampilan berpikir kritis peserta didik pada tiap aspek yang diujikan, diketahui bahwa pada aspek klarifikasi dasar sebagian besar peserta didik termasuk dalam kategori rendah yaitu sebesar 37,50%, ketercapaian pada aspek membangun keterampilan dasar sebagian besar termasuk dalam kategori sangat rendah yaitu sebesar 46,88%, ketercapaian untuk aspek klarifikasi lanjut sebagian besar termasuk dalam kategori sangat rendah yaitu sebesar 56,25%, dan ketercapaian untuk aspek mengatur strategi dan taktik sebagian besar termasuk dalam kategori rendah yaitu sebesar 34,38%. Hasil analisis menunjukkan terdapat beberapa peserta didik dengan kemampuan berpikir kritis yang sangat tinggi, namun di sisi lain ada pula peserta didik dengan kemampuan berpikir kritis yang sangat rendah. Hal ini tentunya dipengaruhi oleh beberapa faktor, salah satunya yaitu faktor karakteristik kelas. Kelas yang digunakan sebagai subjek uji coba merupakan kelas dengan input peserta didik yang memiliki kemampuan bervariasi. Artinya, peserta didik pada kelas tersebut tidak memiliki kemampuan yang sama sedari awal, tetapi terdapat peserta didik dengan kemampuan yang sangat tinggi dan peserta didik dengan kemampuan yang cukup. Meskipun demikian, instrumen keterampilan berpikir kritis yang dikembangkan ini mampu mengukur kemampuan peserta didik dari yang paling tinggi hingga kemampuan yang paling rendah. Hasil ini sesuai dengan penelitian yang dilakukan oleh Burhanudin (2015) yang menyatakan bahwa penilaian otentik dengan tes tertulis cocok digunakan untuk menilai aspek kognitif peserta didik.

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan maka dapat disimpulkan bahwa instrumen penilaian otentik yang dikembangkan telah memenuhi kriteria layak/valid ditinjau dari validitas isi maupun validitas empiris. Instrumen penilaian otentik yang dikembangkan tergolong reliabel dan dapat digunakan sebagai salah satu instrumen alternatif untuk mengukur keterampilan proses sains peserta didik pada aspek menginvestigasi, menginterpretasi, dan membuat model, serta untuk mengukur keterampilan berpikir kritis peserta didik pada aspek klarifikasi dasar, membangun keterampilan dasar, klarifikasi lanjut, dan mengatur strategi/taktik. Berdasarkan penelitian yang dilakukan, peneliti memberikan saran untuk penelitian berikutnya, yaitu: Instrumen penilaian otentik dapat dikembangkan lebih lanjut untuk mengukur keterampilan proses sains pada aspek yang lain. Instrumen penilaian otentik yang dikembangkan sebaiknya tidak sekedar mengukur aspek kognitif peserta didik, tetapi juga dapat dikembangkan dan digunakan untuk mengukur aspek psikomotor peserta didik yang terintegrasi dengan pembelajaran.

DAFTAR PUSTAKA

- Azim, S., Khan, M. (2012). Authentic assessment: An instructional tool to enhance students learning. *Academic Research International*, 2(3): 314-320.
- Berta Apriza dan Ali Mahmudi. (2015). Keefektifan Pendekatan PBL dan Discovery Setting TPS ditinjau dari Prestasi, Kemampuan Berpikir Kritis, dan Kepercayaan Diri Siswa. *Jurnal Pendidikan Matematika dan Sains*. 3(2): 101-110.
- Hambleton, R. K., & Swaminathan, H., (1985). *Item Respon Theory Principles and Application*. Boston: Kluwer. Nijhoff Publishing.
- Istiyono, Edi., Mardapi, D. & Suparno. (2014). *Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PhyTHOTS) Peserta Didik SMA*. Jurnal Penelitian dan Evaluasi Pendidikan. Halaman 1-12.
- Johnson, B. (2007). *Contextual Teaching & Learning*. Bandung: Mizan Media Utama.
- Lambertus. (2009). Pentingnya Melatih Keterampilan Berfikir Kritis dalam Pembelajaran Matematika di SD. *Jurnal forum Kependidikan*. Vol28(2): 136-142.
- Mueller, Jon. (2005). The Authentic Assessment Toolbox: Enhancing Student Learning through Online Faculty Development. *Journal of Online Learning and Teaching*, Vol 1(1).
- Nitko, A.J. (1996). *Educational Assesment of Student*. Englewood Cliffs: Merrill Norris, S.P. dan Ennis, R. 1989. *Evaluating Critical Thinking* (dalam R. J. Schwartz & D. N. Perkins (Eds), *The Practitioners' Guide to Teaching Thinking Series*. Pacific Grove, California: Midwest Publications.
- Yulianti, E, Husna, I.Y, dan Susilowati. (2018). The Role of Inquiry-Based Interactive Demonstration Learning Model on VIII Grade Students' Higher Order Thinking Skill. *Journal of Science Education Research*. 2(1), 35-38.
- Wiggins, G. (1990). Practical Assessment, Research & Evaluation. *A Peer- Reviewed Electronic Journal*, 2(2): 1-3.