



**PENGEMBANGAN INSTRUMEN TES *TWO-TIER MULTIPLE CHOICE* BERBASIS HOTS
PADA MATERI RANGKAIAN ARUS BOLAK-BALIK**

Dewi Risma Winari, Masturi

Jurusan Fisika FMIPA Universitas Negeri Semarang

dewirismawnr@students.unnes.ac.id, masturi@mail.unnes.ac.id

Diterima : Desember 2022. Disetujui : Januari 2023. Dipublikasikan : Februari 2023

ABSTRAK

Penelitian ini dilatarbelakangi oleh data PISA 2018 yang menyebutkan peringkat Indonesia berada pada 74 dari 79 negara. Kemampuan berpikir tingkat tinggi dan kompetensi pengajar pada level rendah menjadi faktor yang menyebabkan Indonesia berada pada peringkat bawah. Sehubungan dengan itu tenaga pengajar memerlukan inovasi salah satunya dengan pengembangan instrumen tes berbasis High Order Thinking Skills (HOTS). Tujuan penelitian ini adalah mengetahui tingkat kelayakan dan keefektifan instrumen tes untuk materi rangkaian arus bolak balik. Metode penelitian ini adalah research and development (R&D) dengan model 4D. Instrumen tes terdiri atas 25 soal berbasis HOTS dengan jumlah item soal kategori sukar berjumlah 3 soal, kategori sedang berjumlah 14 soal, dan kategori mudah berjumlah 8 soal. Analisis data dengan pemodelan Rasch berbantuan Software Ministep 5.2.4.0. Instrumen tes yang dikembangkan dikategorikan reliabel dengan perolehan nilai alpha cronbach uji skala kecil dan uji skala besar masing-masing 0,88 dan 0,76. Analisis data tersebut, menyatakan instrumen HOTS yang dikembangkan dalam kategori cukup dan dapat diaplikasikan untuk mengukur kemampuan berpikir tingkat tinggi mahasiswa.

Kata Kunci: *Two-Tier Multiple-Choice test, HOTS, Rasch model.*

ABSTRACT

The research was motivated by PISA data for 2018 which stated that Indonesia was ranked 74th out of 79 countries. Higher-order thinking skills and low teacher competence are one of the factors. Thus, teachers need innovation, one of which is the development of test instruments based on High Order Thinking Skills (HOTS). The purpose of this study was to determine the level of feasibility and effectiveness of the alternating current circuit material test. This research method is research and development (R&D) with 4D model. The test consisted of 25 HOTS-based questions with 3 questions in the difficult category, 14 questions in medium, and 8 questions in the easy category. Data analysis using Rasch modeling by Ministep 5.2.4.0 Software. The test instrument developed was categorized as reliable with Cronbach's alpha scores on small and large scales tests of 0.88 and 0.76, respectively. Data analysis stated that the HOTS instrument developed was in the sufficient category and could be applied to measure students' higher order thinking skills.

Keywords: *Two-Tier Multiple-Choice test, HOTS, Rasch model.*

PENDAHULUAN

Hasil penelitian yang dilakukan oleh PISA (*Programme for International Student Assessment*) menempatkan bangsa Indonesia pada peringkat bawah dari negara-negara partisipasi dalam program yang dikoordinasi oleh *Organization for Economic Cooperation and Development* (OECD). Tes yang dilakukan setiap tiga tahun sekali memposisikan bangsa Indonesia berada di peringkat 72 dari 77 negara di bidang literasi, peringkat 72 dari 78 negara untuk bidang matematika, sedangkan dalam bidang sains Indonesia berada pada peringkat 70 dari 78 negara yang berpartisipasi. Keadaan tersebut sama halnya pada tahun 2015 dan tahun 2013 (<https://nces.ed.gov/surveys/pisa/>)(<https://pskp.kemdikbud.go.id/>). Salah satu cara untuk dapat meningkatkan kualitas pendidikan adalah dengan diciptakan suatu kurikulum yang dapat dijadikan sebagai pedoman dalam pembelajaran. Di Indonesia sendiri pergantian kurikulum sudah sering dilakukan, hal tersebut dilakukan untuk memenuhi kebutuhan peserta didik. Tercatat dalam sepuluh tahun terakhir, sudah dilakukan dua kali pergantian kurikulum, diantaranya adalah kurikulum 2013 revisi (K-13 revisi) dan kurikulum merdeka yang menjadi pedoman pembelajaran terbaru peserta didik (Wibowo, 2014). Berdasarkan Keputusan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi RI Nomor 56/M/2022 tentang pedoman penerapan kurikulum dalam rangka pemulihan pembelajaran menjelaskan bahwa: perlu adanya pemulihan akibat *learning loss* (ketertinggalan pembelajaran) *pasca* pandemi, sehingga satuan pendidikan dan tenaga pengajar perlu mengembangkan kurikulum yang sesuai dengan kondisi satuan pendidikan, potensi daerah masing-masing, dan kebutuhan peserta didik (Menteri Pendidikan, Kebudayaan, 2022). Kurikulum yang dimaksudkan adalah Kurikulum Merdeka, dengan konsep “Merdeka Belajar” bagi peserta didik. Tujuan dilakukan pengembangan kurikulum ini adalah untuk membantu pemulihan pembelajaran *pasca* adanya pandemi COVID-19, sehingga penggunaan

teknologi dan kebutuhan kompetensi di era sekarang menjadi suatu prioritas (Nugraha, 2022) (Brier & lia dwi jayanti, 2020).

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi merancang sebuah kurikulum untuk dapat menghasilkan sumber daya manusia yang produktif dalam menghadapi perkembangan global dan memiliki kemampuan berpikir kritis sesuai dengan tujuan kurikulum merdeka. Pembelajaran dengan pendekatan ilmiah mengimplementasikan pendekatan *scientific* yang terdiri dari tiga ranah, yaitu: pengetahuan (kognitif), sikap (afektif), dan keterampilan (psikomotorik). Ketiga ranah tersebut diharapkan dapat melahirkan peserta didik yang inovatif, kreatif, dan produktif melalui penguatan sikap, keterampilan, dan pengetahuan yang terintegrasi. Sehubungan hal tersebut, dalam mengembangkan ilmu pengetahuan dan teknologi diperlukan kemampuan menguasai konsep dan prinsip dasar, kemampuan berpikir kritis dan kreatif, serta sikap percaya diri yang tinggi bagi masing-masing individu (<https://www.kemdikbud.go.id>).

Faktor yang menyebabkan rendahnya peringkat Indonesia dalam PISA salah satunya adalah peserta didik belum terbiasa menghadapi soal HOTS (*High Order Thinking Skills*). Perlu diberikan soal-soal dengan level *HOTS* agar peserta didik dapat terlatih mengembangkan kemampuan kognitif. Berdasarkan penelitian yang telah dilakukan oleh Edi Istiyono (2014) menyebutkan bahwa instrumen Soal *HOTS* yang diujikan kepada peserta didik dalam kategori rendah, yaitu 78,87%, sedangkan kemampuan peserta didik dalam mengerjakan soal *HOTS* fisika kategori tinggi sebesar 20,94% dan sangat tinggi 0,19%. Penelitian lain yang dilakukan oleh Nursa'adah et al., (2016:25) yang menyatakan bahwa peserta didik yang mampu menyelesaikan soal *HOTS* dalam kategori rendah sebesar 12, dan untuk level sedang dapat di selesaikan oleh 18 peserta didik. Berdasarkan latar belakang di atas, peneliti mencoba untuk menyusun instrumen tes berbasis *HOTS* pada materi

rangkaian arus bolak balik untuk mengetahui kemampuan berpikir kritis mahasiswa.

Higher Order Thinking Skills (HOTS) merupakan proses berpikir yang mengharuskan peserta didik untuk meningkatkan kreativitas dan mengembangkan ide-ide dengan cara yang berbeda. Berpikir tingkat tinggi merupakan kemampuan berpikir yang tidak sekadar mengingat, menuliskan kembali, dan merujuk tanpa mengolah informasi, namun kemampuan untuk menelaah permasalahan secara kritis, kreatif, kreasi, dan tentunya peserta didik dapat memecahkan masalah tersebut. Indikator kemampuan berpikir tingkat tinggi menurut Anderson & Krathwohl adalah (1) kemampuan menganalisis yang terdiri dari kemampuan membedakan sebab akibat sebuah permasalahan dan mengidentifikasi suatu pernyataan, (2) kemampuan mengevaluasi yang meliputi membuat hipotesis, dan melakukan pengujian, serta (3) kemampuan mengkreasi yang berupa kegiatan merancang suatu percobaan sederhana untuk menyelesaikan suatu permasalahan (Mariani, R., Ansori, H., & Mawaddah, S., 2021).

Indikator HOTS yang demikian masih minim didapatkan peserta didik. Salah satu cara yang dapat dilakukan untuk mengembangkan kemampuan tersebut adalah melalui instrumen soal *two-tier multiple choice* berbasis HOTS. Tes yang sering ditempuh oleh peserta didik adalah model pilihan ganda biasa, dan yang membedakan instrumen ini dengan soal pilihan ganda biasanya adalah terdapat *tier* kedua yang berisi alasan mengapa peserta didik memilih jawaban tersebut pada *tier* pertama. Keunggulan tes *two-tier multiple choice* yaitu dapat mengungkap alasan dibalik opsi yang dipilih peserta didik pada tahap pertama, hal ini secara tidak langsung mengurangi tingkat kesalahan peserta didik yaitu kemungkinan menjawab dengan metode menebak. Peserta didik mendapatkan skor sempurna apabila pada kedua *tier* menjawab dengan benar (Lengkong et al., 2021) (Ivanjek et al., 2021) (Fратиwi et al., 2017). Tes *two-ter multiple choice* dapat digunakan untuk mengidentifikasi dan

mengevaluasi konsepsi peserta didik dalam bidang tertentu. Selain itu, tes diagnostik pilihan ganda dua tingkat memiliki dua keunggulan lainnya dibandingkan pertanyaan pilihan ganda *one-tier* konvensional.

METODE PENELITIAN

Jenis penelitian menggunakan pengembangan R&D model 4D, dengan tahapan meliputi : (1) pendahuluan (*define*), (2) perancangan (*design*), (3) pengembangan (*develop*), (4) penyebaran (*disseminate*) (Lengkong et al., 2021) (Fратиwi et al., 2017) (Mandagi et al., 2021). Subjek dalam penelitian ini adalah 50 mahasiswa UNNES dan UIN Walisongo Semarang dengan jumlah soal yang digunakan 25 butir soal *two-tier multiple choice* (Ivanjek et al., 2021) (Mešić et al., 2019). Pedoman penskoran yang digunakan dalam penelitian ini adalah menggunakan *Graded Response Models* (GRM), dengan penskoran dilakukan di tiap kategori pada item tes yang disusun secara berurutan (Sari & Andriani, 2019). Setiap soal, peserta didik harus menjawab dua tingkatan untuk setiap butir soal dengan perolehan *perfect score* adalah 2 (Wardani et al., 2015; Uccio et al., 2019; Fратиwi et al., 2017). Tahap *define* dilakukan dengan studi literatur mengenai kemampuan berpikir kritis peserta didik sedangkan tahapan *design* adalah menyusun instrumen soal sampai pedoman penskoran. Apabila responden menjawab benar di salah satu *tier*, maka hanya mendapatkan score 1, sedangkan apabila responden menjawab salah untuk kedua *tier* atau tidak menjawab maka mendapatkan score 0.

Tahap *develop* melakukan uji validitas dengan tujuan untuk mengetahui keefektivitasan instrumen tes sebagai alternatif penilaian dalam mengukur kemampuan berpikir kritis peserta didik. Tahap pertama adalah validasi soal yang dilakukan oleh para ahli. Validator pada penelitian ini adalah dua dosen fisika Universitas Negeri Semarang yang merupakan rekomendasi dari dosen pembimbing (Ilmi et al., 2020; Yuliatun et al., 2020). Analisis data yang digunakan adalah pemodelan *Rasch* berbantuan *software*

Ministep 5.2.4.0 dengan panduan kriteria (Boone *et al.*, 2014) (Wati *et al.*, 2019):

- (1) Nilai *Outfit Mean Square (MNSQ)* yang diterima: $0,5 < MNSQ < 1,5$;
- (2) Nilai *Outfit ZStandard (ZSTD)* yang diterima: $-2,0 < ZSTD < +2,0$;
- (3) Nilai *Point Measure Correlation (Pt Measure Corr)* yang baik: $0,4 < Pt Measure Corr < 0,85$.

Analisis reliabilitas berfungsi untuk mengetahui tingkat konsistensi dari instrument mencakup nilai *Alpha Cronbach*, nilai *Person Reliability* dan *Item Reliability* (Sumintono dan Widhiarso, 2015).

HASIL DAN PEMBAHASAN

Analisis butir soal dilakukan kepada validator ahli yang merupakan dosen UNNES dan kepada 30 mahasiswa didapatkan nilai logit outfit MNSQ: $0,63 \leq \text{outfit MNSQ} \leq 2,26$; nilai logit outfit ZSTD: $-1,85 \leq \text{outfit ZSTD} \leq 1,59$; nilai logit PT. Measure Corr. bernilai $0,15 \leq \text{PT. Measure Corr} \leq 0,74$. Analisis butir soal dengan Ministep 5.2.4.0 Software dapat dilihat pada Gambar 1 berikut.

TABLE 10.1 C:\Users\Lenovo\Downloads\2_data skal ZOU499HS.TXT Jul 6 2022 0: 5
 INPUT: 30 Person 25 Item REPORTED: 30 Person 25 Item 2 CATS MINISTEP 5.2.4.0
 Person: REAL SEP.: 1.83 REL.: .77 ... Item: REAL SEP.: 1.47 REL.: .68

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT MNSQ	PTMEASUR-AL ZSTD	EXACT MATCH OBS% EXP%	Item			
15	5	30	.86	.56	1.37	1.15	2.26	1.59	A .19 .42	78.6 84.3	A15	
22	5	30	.86	.56	1.37	1.15	2.26	1.59	B .19 .42	78.6 84.3	A22	
9	3	30	1.60	.67	1.19	.56	2.23	1.24	C .19 .35	85.7 89.2	A9	
6	4	30	1.28	.60	1.45	1.22	1.91	1.15	D .15 .39	82.1 86.4	A6	
20	4	30	1.28	.60	1.45	1.22	1.91	1.15	E .15 .39	82.1 86.4	A20	
14	7	30	.31	.50	1.36	1.33	1.61	1.46	F .27 .47	67.9 79.8	A14	
23	7	30	.31	.50	1.36	1.33	1.61	1.46	G .27 .47	67.9 79.8	A23	
1	9	30	-1.16	.47	1.13	.64	1.68	1.59	H .39 .51	78.6 76.1	A1	
4	9	30	-1.16	.47	1.13	.64	1.68	1.59	I .39 .51	78.6 76.1	A4	
18	8	30	.07	.48	1.20	.86	1.13	.43	J .48 .49	75.0 77.7	A18	
8	8	30	.07	.48	1.04	.26	.97	.07	K .47 .49	82.1 77.7	A8	
12	5	30	.86	.56	.96	-.04	.63	-.40	L .47 .42	85.7 84.3	A12	
25	5	30	.86	.56	.96	-.04	.63	-.40	M .47 .42	85.7 84.3	A25	
16	18	30	-1.97	.45	.95	-.20	.85	-.26	N .59 .56	78.6 74.1	A16	
21	18	30	-1.97	.45	.95	-.20	.85	-.26	O .59 .56	78.6 74.1	A21	
19	6	30	.57	.52	.80	-.65	.74	-.31	P .53 .45	89.3 82.1	A19	
13	14	30	-1.18	.44	.78	-1.11	.66	-1.22	Q .67 .55	75.0 73.8	A13	
24	14	30	-1.18	.44	.78	-1.11	.66	-1.22	R .67 .55	75.0 73.8	A24	
10	7	30	.31	.50	.77	-.89	.73	-.42	S .57 .47	89.3 79.8	A10	
3	8	30	.07	.48	.76	-1.03	.54	-1.12	T .63 .49	82.1 77.7	A3	
7	8	30	.07	.48	.76	-1.03	.54	-1.12	U .63 .49	82.1 77.7	A7	
2	8	30	.07	.48	.67	-1.50	.48	-1.34	V .67 .49	89.3 77.7	A2	
11	8	30	.07	.48	.67	-1.50	.48	-1.34	W .67 .49	89.3 77.7	A11	
5	15	30	-1.37	.44	.62	-2.13	.52	-1.85	X .75 .56	85.7 74.0	A5	
17	15	30	-1.37	.44	.62	-2.13	.52	-1.85	Y .75 .56	85.7 74.0	A17	
MEAN	8.7	30.0	.00	.51	1.00	-.13	1.14	.01		81.1	79.3	
P.S.D	4.3	.0	.97	.06	.27	1.10	.64	1.19		5.9	4.5	

TABLE 10.3 C:\Users\Lenovo\Downloads\2_data skal ZOU499HS.TXT Jul 6 2022 0: 5
 INPUT: 30 Person 25 Item REPORTED: 30 Person 25 Item 2 CATS MINISTEP 5.2.4.0

Gambar 1. Hasil Uji Coba Validasi Butir Soal

Gambar 1 merupakan tampilan analisis data menggunakan rasch model untuk uji validasi butir soal. Hasil uji coba instrumen dilakukan kepada mahasiswa pendidikan fisika

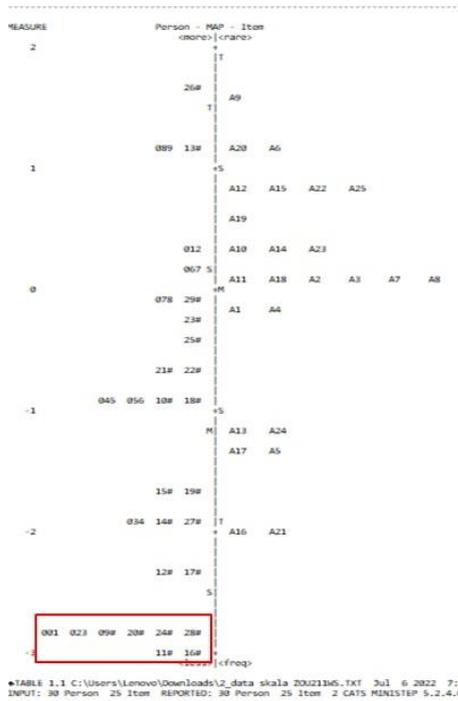
UNNES semester 5 menyatakan dari 25 soal yang dinyatakan valid sebanyak 19 soal, sedangkan 6 soal dinyatakan kurang valid. Data gambar 1 menunjukkan terdapat beberapa soal dengan keterangan tidak diterima, namun dinyatakan valid dan kurang valid. Sebagai contoh adalah kode soal A11 masuk dalam kategori valid, dikarenakan keterangan pada Outfit MNSQ dan Pt. Measure Corr. diterima. Soal dikategorikan fit atau baik apabila memenuhi minimal dua syarat dari ketiga kriteria panduan penilaian menurut Boone. Suatu soal apabila hanya memenuhi satu syarat kriteria maka dinyatakan tidak valid (Tiruneh *et al.*, 2017; Ilmi *et al.*, 2020; Yuliatun *et al.*, 2020). Faktor yang menyebabkan suatu item soal tidak valid yaitu karena banyaknya soal yang dianggap sukar oleh peserta didik, yaitu item soal dengan tipe perhitungan, dan penalaran, serta terdapat pembatasan waktu dalam pengerjaan soal, responden kurang memahami betul mengenai materi yang diujikan, sehingga dalam memilih jawaban relatif cari aman dengan memilih jawaban ragu-ragu (Tiruneh *et al.*, 2017).

Tingkat kesukaran soal digunakan untuk mengetahui persebaran keseluruhan soal yang terdiri dari soal kategori mudah, sedang dan sukar, soal yang baik adalah soal yang tidak terlalu mudah dan tidak terlalu sukar. Tingkat kesukaran soal berdasarkan pemodelan Rasch ditunjukkan dengan nilai logit. Semakin tinggi nilai logit (measure) menunjukkan item soal tersebut memiliki tingkat kesukaran yang tinggi. Berdasarkan analisis yang dilakukan, tingkat kesukaran item soal adalah $-1,97 \leq \text{measure} \leq 1,60$, sehingga secara keseluruhan instrumen soal yang disusun memiliki komposisi yang baik untuk setiap kategori soalnya. Persebaran tingkat kesukaran butir soal dapat dilihat pada Tabel 1.

Tabel 1. Persebaran Kesukaran Butir Soal

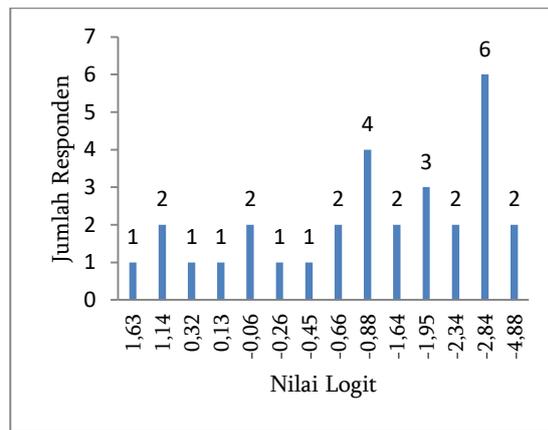
Kategori item soal	Jumlah butir soal	Presentase
Mudah	8	32,00%
Sedang	14	56,00%
Sukar	3	12,00%
Jumlah	25	100,00%

Tabel 1. menunjukkan bahwa, jumlah item soal kategori sukar berjumlah 3 soal, kategori sedang berjumlah 14 soal, dan kategori mudah berjumlah 8 soal. Gambar 2 adalah hasil analisis dengan menggunakan permodelan *Rasch*.



Gambar 2. Persebaran Peserta Didik dalam Menjawab Soal

Gambar 2 menunjukkan persebaran peserta didik dalam menjawab soal menurut analisis *Rasch Model*. Bagian sebelah kiri garis putus-putus menunjukkan persebaran kemampuan mahasiswa sedangkan sebelah kanan menunjukkan persebaran item soal. Responden dengan nomor 26 adalah yang paling pintar karena berada pada posisi yang paling tinggi, sedangkan untuk item soal yang paling sukar adalah soal dengan kode A9, kemudian soal dengan kode A6 dan A20. Mahasiswa dengan nomor urut 1, 23, 9, 20, dan 24 adalah mahasiswa dengan kemampuan berpikir kritis yang masih rendah. Hal tersebut disebabkan karena jumlah item soal yang dikerjakan dengan jawaban salah lebih banyak dibandingkan dengan item soal yang jawabannya benar (Susac et al., 2018). Perolehan skor *logit* mahasiswa dapat dilihat pada Gambar 3.



Gambar 3. Perolehan Skor *Logit*

Gambar 3 menunjukkan perolehan skor logit pada uji skala kecil. Rata-rata skor logit sebesar $-1,40$ dari skor maksimum 2, sehingga dapat dikatakan bahwa kemampuan berpikir tingkat tinggi mahasiswa pendidikan fisika UNNES dalam kategori “rendah”. Analisis reliabilitas berfungsi untuk mengetahui tingkat konsistensi dari instrumen soal ketika diujikan kepada beberapa subjek penelitian. Uji reliabilitas diperoleh dengan menggunakan *rasch model* mencakup nilai *Alpha Cronbach*, nilai *Person Reliability* dan *Item Reliability* (Sumintono dan Widhiarso, 2015; Wati et al., 2019). Nilai kriteria reliabilitas soal dapat dilihat pada Tabel 2.

Tabel 2. Kriteria Reliabilitas Butir Soal

Nilai	Kriteria
$< 0,5$	Buruk
$0,5 - 0,6$	Jelek
$0,6 - 0,7$	Cukup
$0,7 - 0,8$	Bagus
$> 0,8$	Bagus Sekali

Tabel 2 menunjukkan kriteria reliabilitas butir soal menurut Sumintono dan Widhiarso (2015). Berdasarkan kriteria diatas, hasil penelitian diperoleh nilai *Alpha Cronbach* sebesar 0,88 termasuk pada kategori bagus, nilai *Person Reliability* sebesar 0,77 kategori cukup dan *Item Reliability* sebesar 0,68 kategori cukup (Marfu'i et al., 2019). Hasil Rekapitulasi analisis reliabilitas butir soal dapat dilihat pada Tabel 3.

Tabel 3. Rekapitulasi Analisis Reliabilitas

Keterangan	Logit	Kriteria
Person Reliability	0,77	Bagus
Item Reliability	0,68	Cukup
Alpha Cronbach	0,88	Reliabel

Tabel 3 menunjukkan nilai *alpha Cronbach* yang digunakan untuk mengukur interaksi antara responden dan item soal menunjukkan bagus atau reliabel dengan nilai logit 0,88 (Kane et al., 2016). Hal ini menunjukkan adanya kesesuaian antara responden dengan instrumen yang digunakan (Sumintono & Widhiarso, 2015). Berdasarkan hasil uji reliabilitas tersebut dapat disimpulkan bahwa instrumen yang dikembangkan reliabel untuk dikembangkan di lapangan sesuai dengan perolehan *logit person reliability* dan *item reliability*.

Pengambilan data uji skala besar dilakukan dengan menggunakan link *google form* pada dua universitas, yaitu Universitas Negeri Semarang dan UIN Walisongo Semarang dengan jumlah responden sebanyak 50 mahasiswa, dengan jumlah soal yang diujikan sebanyak 25 item soal *two-tier multiple choice*. Analisis data yang telah dilakukan, diperoleh rincian hasil nilai *logit outfit MNSQ* $0,19 \leq outfit MNSQ \leq 3,39$; nilai *logit outfit ZSTD* $-1,78 \leq outfit ZSTD \leq 1,70$; nilai *logit PT. Measure Corr.* bernilai $-0,01 \leq PT. Measure Corr \leq 0,65$. Pengukuran kemampuan berpikir tingkat tinggi (θ) mengacu pada perolehan nilai *logit* pada *tools item output tabel 17* yang terdapat dalam *Software Ministep 5.2.4.0*. Pada pendekatan *Rasch Model* kemampuan dikuantifikasikan dalam bentuk *logit* yang memiliki rentang -4 s.d. $+4$ (Sumintono & Widhiarso, 2015). Interval perolehan skor logit dalam mengukur kemampuan berpikir kritis dapat dilihat pada Tabel 4.

Tabel 4. Kategori Kemampuan Berpikir Kritis

Skor logit	kategori
$0,67 < \theta \leq 2$	Tinggi
$-0,67 < \theta \leq 0,67$	Sedang
$-2 < \theta \leq -0,67$	Rendah

Tabel 4 menunjukkan interval perolehan nilai *logit* dan kategori kemampuan berpikir kritis menurut Boone et al., 2014. Hasil analisis

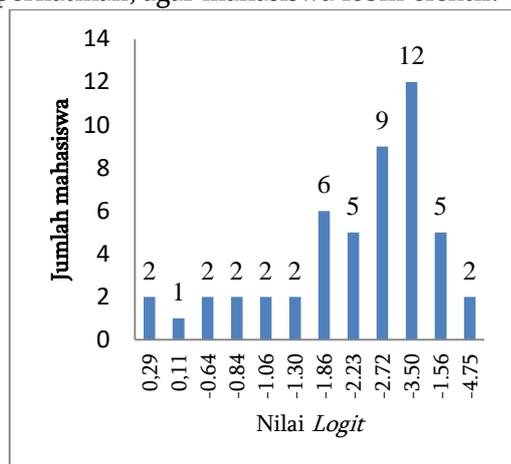
dengan menggunakan Rasch model dapat dilihat pada Gambar 4 di bawah ini.

Gambar 4. Perolehan *Logit* Mahasiswa

Gambar 4 merupakan perolehan *logit* berdasarkan instrumen yang diujikan dalam skala besar pada mahasiswa Fisika Universitas Negeri Semarang dan UIN Walisongo Semarang dengan nilai *logit outfit MNSQ* $0,19 \leq outfit MNSQ \leq 3,39$; nilai *logit outfit ZSTD* $-1,78 \leq outfit ZSTD \leq 1,70$; nilai *logit PT. Measure Corr.* bernilai $-0,01 \leq PT. Measure Corr \leq 0,65$. Berdasarkan Gambar 4 rata-rata nilai logit pada uji skala besar adalah $-2,26$. Pada perolehan nilai rata-rata *logit* tersebut, maka instrumen tes yang digunakan dalam kategori "rendah" sesuai kriteria pada tabel 4, hal tersebut dikarenakan banyak mahasiswa yang belum menjawab benar untuk masing-masing item soalnya (Planinic et al., 2019).

Hal yang dapat dilakukan adalah melakukan kegiatan *pre-test*, dan membuat instrumen soal yang tidak terlalu sukar maupun tidak terlalu mudah namun dapat mengetahui tingkat kemampuan berpikir kritis mahasiswa. Apabila soal tersebut terlalu sulit maka mahasiswa akan lebih memilih asal sehingga presentase menjawab dengan benar sangat kecil. Selain itu mahasiswa harus sudah mendapatkan materi rangkaian arus bolak-balik, baik materi tersebut disampaikan oleh tenaga pengajar maupun materi tersebut

disampaikan oleh peneliti. Waktu yang digunakan untuk pengerjaan harus lebih diperhatikan, agar mahasiswa lebih efektif.



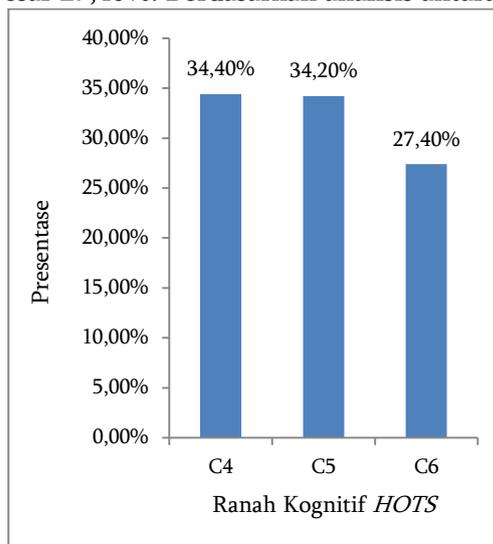
Gambar 5. Persebaran Nilai *Logit* Mahasiswa

Persebaran nilai *logit* untuk uji skala besar dapat dilihat pada Gambar 5. Perolehan nilai sebesar -2,26. Dengan demikian instrumen tes yang digunakan untuk mengukur kemampuan berpikir tingkat tinggi peserta didik pada mahasiswa semester 2 dikategorikan "rendah" pada materi arus bolak balik. Hal tersebut disebabkan karena mahasiswa tidak diberikan *pre-test* untuk mengukur kemampuan awal mahasiswa dan bahan ajar untuk materi yang hendak diujikan.

Penelitian yang dilakukan adalah mengukur kemampuan berpikir kritis sesuai Taksonomi Bloom revisi Anderson. Kriteria instrumen soal *HOTS* yang dianalisis dalam penelitian ini ranah taksonomi Bloom C4, C5 dan C6. Ranah C4 atau kemampuan menganalisis informasi dari butir soal yang berupa diagram, ataupun soal cerita. Ranah C5 merupakan kegiatan mengevaluasi untuk menentukan informasi atau nilai suatu benda berdasarkan kriteria tertentu, sedangkan ranah C6 adalah proses mengkreasi dengan tujuan untuk merancang dan mengembangkan ide atau gagasan yang baru.

Gambar 6 menunjukkan jumlah *presentase* mahasiswa yang dapat menjawab benar untuk masing-masing ranah kognitif pada materi rangkaian arus bolak-balik. Dapat dilihat bahwa mahasiswa yang menjawab benar pada ranah C4 (menganalisis) sebesar 34,40%, 32,40% untuk ranah C5 (mengevaluasi), dan

ranah C6 (mengkreasi) memperoleh *presentase* sebesar 27,40%. Berdasarkan analisis antara



Gambar 6. Presentase Mahasiswa yang Menjawab Benar

KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa instrumen tes berbasis *HOTS* dengan pedoman penskoran *Graded Respons Models (GRM)* pada materi rangkaian arus bolak balik yang dikembangkan telah memenuhi standar kelayakan instrumen yang valid dan reliabel. Uji validitas butir soal oleh ahli skor rata-rata untuk ketiga aspek (substansi, konstruksi, dan bahasa) sebesar 93,06%. Kemudian hasil uji reliabilitas pada tahap uji skala kecil diperoleh nilai *alpha cronbach* 0,88 dan masuk dalam kategori reliabel, dengan *logit person reliability* 0,77 (bagus); *logit item reliability* sebesar 0,68 (cukup), sedangkan pada tahap uji skala besar nilai *alpha cronbach* sebesar 0,76 dengan perolehan *logit person reliability* 0,50 (jelek) dan *logit item reliability* 0,62 (cukup). Oleh sebab itu, instrumen *HOTS* yang disusun dalam kategori cukup dapat diaplikasikan untuk mengukur kemampuan berpikir tingkat tinggi mahasiswa.

UCAPAN TERIMA KASIH

Penghargaan disampaikan kepada Jurusan Pendidikan Fisika UIN Walisongo Semarang yang sudah memfasilitasi penelitian ini.

DAFTAR PUSTAKA

- Boone, W.J., et al. 2014. Rasch Analysis in The Human Sciences. Dordrecht: Springer
- Brier, J., & lia dwi jayanti. (2020). Pengembangan Kurikulum Merdeka. (Vol. 21, Issue 1).
- Brier, J., & lia dwi jayanti. (2020). *Pengembangan Kurikulum Merdeka*. Vol. 21, Issue 1. <http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203>
- Fратиwi, N. J., Kaniawati, I., Suhendi, E., Suyana, I., & Samsudin, A. (2017). The transformation of two-tier test into four-tier test on Newton's laws concepts. *AIP Conference Proceedings*, 1848(May 2017). <https://doi.org/10.1063/1.4983967>
- Ilmi, A. M., Sukarmin, & Sunarno, W. (2020). Development of TPACK based-physics learning media to improve HOTS and scientific attitude. *Journal of Physics: Conference Series*, 1440(1). <https://doi.org/10.1088/1742-6596/1440/1/012049>
- Ivanjek, L., Morris, L., Schubatzky, T., Hopf, M., Burde, J. P., Haagen-Schützenhöfer, C., Dopatka, L., Spatz, V., & Wilhelm, T. (2021). Development of a two-tier instrument on simple electric circuits. *Physical Review Physics Education Research*, 17(2), 20123. <https://doi.org/10.1103/PhysRevPhysEducRes.17.020123>
- Kane, S. N., Mishra, A., & Dutta, A. K. (2016). Preface: International Conference on Recent Trends in Physics (ICRTP 2016). *Journal of Physics: Conference Series*, 755(1). <https://doi.org/10.1088/1742-6596/755/1/011001>
- Lengkong, M., Istiyono, E., Rampean, B. A. O., Tumanggor, A. M. R., & Nirmala, M. F. T. (2021). Development of Two-Tier Test Instruments to Detect Student's Physics Misconception. *Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, 528(Icриems 2020), 561–566. <https://doi.org/10.2991/assehr.k.210305.082>
- Mandagi, A. F., Iswanto, B. H., & Sugihartono, I. (2021). Virtual Microscopic Simulation (VMS) for physics learning of the photoelectric effect in high school. *Journal of Physics: Conference Series*, 2019(1). <https://doi.org/10.1088/1742-6596/2019/1/012013>
- Marfu'i, L. N. R., Ilfiandra, & Nurhuda. (2019). The analysis of critical thinking skills test in social-problems for physics education students with Rasch Model. *Journal of Physics: Conference Series*, 1280(5). <https://doi.org/10.1088/1742-6596/1280/5/052012>
- Menteri Pendidikan, Kebudayaan, R. dan T. (2022). *Salinan Keputusan Menteri Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia Nomor 56/M/2022 tentang Pedoman Penerapan Kurikulum dalam Rangka Pemulihan Pembelajaran*.
- Mariani, R., Ansori, H., & Mawaddah, S. (2021). Kemampuan Berpikir Tingkat Tinggi Menurut Teori Anderson dan Krathwohl pada Siswa SMP Kelas IX. *Jurmadikta*, 1 (1): 49-55.
- Mešić, V., Neumann, K., Aviani, I., Hasović, E., Boone, W. J., Erceg, N., Grubelnik, V., Sušac, A., Glamočić, D. S., Karuza, M., Vidak, A., Alihodžić, A., & Repnik, R. (2019). Measuring students' conceptual understanding of wave optics: A Rasch modeling approach. *Physical Review Physics Education Research*, 15(1), 1–20.

- <https://doi.org/10.1103/PhysRevPhysEducRes.15.010115>
- Nugraha, T. S. (2022). *Inovasi Kurikulum*. 250–261.
- Nursa'adah, E., Kurniawati, D., & Yunita. (2016). Analisis Kemampuan Kognitif Mahasiswa Pada Konsep Asam-Basa Menggunakan Tes Berdasarkan Taksonomi Bloom Revisi. *EduChemia*, *1*(1), 25–35.
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 20111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Sari, R. N., & Andriani, S. (2019). Penggunaan Graded Response Models (Grm) Dalam. *Prosiding Seminar ...*, 175–188. <http://ejournal.radenintan.ac.id/index.php/pspm/article/view/4307>
- Scotti di Uccio, U., Colantonio, A., Galano, S., Marzoli, I., Trani, F., & Testa, I. (2019). Design and validation of a two-tier questionnaire on basic aspects in quantum mechanics. *Physical Review Physics Education Research*, *15*(1), 10137. <https://doi.org/10.1103/physrevphyseducres.15.010137>
- Susac, A., Planinic, M., Klemencic, D., & Milin Sipus, Z. (2018). Using the Rasch model to analyze the test of understanding of vectors. *Physical Review Physics Education Research*, *14*(2), 23101. <https://doi.org/10.1103/PhysRevPhysEducRes.14.023101>
- Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, *15*(4), 663–682. <https://doi.org/10.1007/s10763-016-9723-0>
- Wardani, R., Yamtinah, S., & Mulyani, B. (2015). Instrumen Penilaian Two-Tier Test Aspek Pengetahuan Untuk Mengukur Keterampilan Proses Sains (Kps) Pada Pembelajaran Kimia Untuk Siswa Sma/Ma Kelas X. *Jurnal Pendidikan Kimia Universitas Sebelas Maret*, *4*(4), 156–162.
- Wati, M., Mahtari, S., Hartini, S., & Amalia, H. (2019). A Rasch model analysis on junior high school students' scientific reasoning ability. *International Journal of Interactive Mobile Technologies*, *13*(7), 141–149. <https://doi.org/10.3991/ijim.v13i07.10760>
- Wibowo, H. (2014). Perubahan kurikulum di indonesia: studi kritis tentang upaya menemukan kurikulum pendidikan islam yang ideal. *Raudhah*, *IV*(1), 49–70.
- Yuliatun, T., Jumadi, Istiyono, E., & Putranta, H. (2020). Development Of Physics-Two Tier Test Based On Jumanji Game To Measure Conceptual Understanding Of Work And Energy. *International Journal of Scientific & Technology Research*, *9*(2), 3888–3895. www.ijstr.org