



## PENGEMBANGAN DAN KARAKTERISASI TES KETERAMPILAN BERPIKIR KRITIS MATERI TEKANAN BERDASARKAN TEORI RESPON BUTIR

<sup>1</sup>Revi Novia, <sup>1</sup>Taufik Ramlan Ramalis, <sup>1</sup>Ridwan Efendi

<sup>1</sup>Program Studi Pendidikan Fisika, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi 229 Bandung 40154, Indonesia

\* E-mail: [reviinoviaa@gmail.com](mailto:reviinoviaa@gmail.com)

### ABSTRAK

Keterampilan berpikir kritis merupakan keterampilan yang harus dimiliki siswa pada abad 21. Dalam dunia pendidikan, keterampilan berpikir kritis siswa dapat diketahui melalui penilaian menggunakan alat ukur keterampilan berpikir kritis. Namun pada kenyataannya masih jarang sekali pendidik menyusun alat ukur yang dapat merangsang siswa untuk berpikir kritis. Alasan utamanya dikarenakan sulitnya mengkonstruksi dan menganalisis alat ukur keterampilan berpikir kritis karena membutuhkan waktunya yang cukup lama. Penelitian ini bertujuan untuk mengembangkan dan mengkarakterisasi tes keterampilan berpikir kritis berdasarkan teori respon butir. Metode penelitian yang digunakan adalah metode pengembangan. Tes dikonstruksi berdasarkan 16 *subskill* dari 6 aspek keterampilan berpikir kritis pada materi tekanan. Partisipan dalam penelitian ini yaitu 151 siswa di enam SMP di Lembang. Analisis data dalam penelitian ini berdasarkan teori respon butir dan dibantu dengan *software* IRTPRO4.2 *Student*. Hasil analisis dari kurva karakteristik tes menunjukkan bahwa parameter kesukaran ( $b_{jk}$ ) terentang antara -0,15 sampai 0,92 dengan indeks kesukaran termasuk dalam kategori sedang dan parameter daya pembeda ( $a$ ) bernilai 0,49 yang berarti tes memiliki daya pembeda yang baik. Perpotongan antara total fungsi informasi dan SEM (standard error measurement) berada pada rentang -1,10 sampai 1,7. Hal ini berarti, keseluruhan tes ini dapat digunakan untuk mengukur kemampuan berpikir kritis dari tingkat kemampuan rendah (dibawah rata-rata) sampai kemampuan tinggi (diatas rata-rata).

Kata kunci: keterampilan berpikir kritis, pengembangan dan karakterisasi, teori respon butir

### ABSTRACT

Critical thinking skills are the skills students must have in the 21st century. In the world of education, students' critical thinking ability can be seen through assessment using the critical thinking skill test. But in reality it is rare for educators to prepare a measurement tool that can stimulate students to think critically. The main reason is because it is difficult to build and analyze critical thinking skills because it takes a long time. This study aims to develop and characterize critical thinking skills tests based on item response theory. The research method used is a method of development. The test is built on the basis of 16 subskill of 6 aspects of critical thinking skills in the pressure material. Participants in this study were 151 students in six middle schools in Lembang. The analysis of the data of this research is based on item response theory and assisted by using *software* IRTPRO4.2 *Student*. The result of analysis from characteristic curve of the tests shows that the parameter of level of difficulty ( $b$ ) between -0.15 to 0.92 with index of level of difficulty included in the medium category and parameter of discriminating power ( $a$ ) value of 0.49 which means that the test has a good category. The information function amount 18.83 and the intersection between the total information function and SEM (standard measurement error) is in the range -1.10 to 1.7. This means, the test is valid and it can be used to measure the critical thinking ability of low levels (below average) until the high ability levels (above average).

Keywords: critical thinking skill, development and characterization, item response theory

## PENDAHULUAN

Seiring dengan perkembangan pengetahuan dan teknologi sekarang ini dibutuhkan sumber daya manusia yang mampu berpikir kritis untuk menyelesaikan permasalahan yang kompleks. Hal ini sejalan dengan pendapat Beybee dan Fuchs bahwa pendidikan ilmu pengetahuan dan teknologi harus dilihat sebagai hal yang penting untuk mencapai kompetensi tenaga kerja yang diinginkan yang meliputi berpikir kritis, keterampilan komunikasi yang kompleks dan kemampuan untuk memecahkan masalah semi-terstruktur (Beybee & Fuchs, 2006, hlm. 350). Sehingga berpikir kritis harus sudah mulai diajarkan dan dibiasakan dalam pembelajaran di sekolah. Dalam kurikulum pendidikan tertuang peserta didik diharapkan mampu mengembangkan keterampilan berpikir kritis sehingga peserta didik memiliki *softskill* yang menunjang kehidupan dimasa yang akan datang. Sehubungan dengan adanya tuntutan peningkatan kualitas sumber daya manusia yang intelektual dan berpikir kritis melalui pendidikan, maka perlu dilakukan peningkatan penguasaan pengetahuan pada berbagai mata pelajaran salah satunya fisika. Kemampuan berpikir kritis yang dimiliki oleh seseorang tidak dapat dimiliki secara langsung melainkan diperoleh melalui latihan (Rofiah, dkk., 2013, hlm. 18)

Oleh karena itu, tuntutan keterampilan tersebut hendaknya menjadi perhatian bagi guru agar keterampilan tersebut dapat dilatihkan dan dikembangkan selama proses pembelajaran di kelas. Strategi pembelajaran yang didukung secara empiris yang mendorong, merangsang, dan memfasilitasi kemampuan siswa serta transfer keterampilan berpikir kritis suatu hal yang penting untuk mengembangkan keterampilan berpikir kritis (Tiruneh, dkk., 2014, hlm. 3). Berpikir kritis adalah berpikir yang memiliki tujuan (pembuktian suatu hal, menafsirkan maksud dari sesuatu, dan pemecahan masalah), tetapi berpikir kritis bisa melakukan kolaborasi, usaha yang tidak bersaing (Facione, 2013, hlm. 4). Ada beberapa kategori indikator menurut Facione yaitu interpretasi, analisis, interferensi, evaluasi, eksplanasi, dan regulasi diri.

Menafsirkan (*interpretation*) merupakan kemampuan seseorang untuk memahami dan mengekspresikan maksud dari suatu situasi, data, penilaian, aturan, prosedur, atau kriteria yang bervariasi. *Subskillnya* meliputi meliputi kategorisasi, pemecahan kode, dan klarifikasi makna. Menganalisis (*Analysis*) merupakan kemampuan seseorang untuk mengklarifikasi kesimpulan berdasarkan hubungan antara informasi serta konsep, dengan pertanyaan yang ada dalam masalah. *Subskillnya* meliputi memeriksa ide-ide, mengidentifikasi argument, dan menganalisis argument.

Mengevaluasi (*evaluation*) merupakan kemampuan seseorang untuk menilai kredibilitas dari suatu pernyataan atau representasi lain dari pendapat seseorang atau menilai suatu kesimpulan berdasarkan hubungan antara informasi dan konsep, dengan pertanyaan yang ada dalam suatu masalah. *Subskillnya* meliputi penilaian klaim dan penilaian argument. Menyimpulkan (*inference*) merupakan kemampuan seseorang untuk mengidentifikasi elemen-elemen yang dibutuhkan dalam membuat kesimpulan yang rasional, dengan mempertimbangkan informasi-informasi yang relevan dengan suatu masalah dan konsekuensinya berdasarkan data yang ada. *Subskillnya* meliputi mengumpulkan informasi, menduga alternatif, dan membuat kesimpulan.

Menjelaskan (*explanation*) merupakan kemampuan seseorang untuk menyatakan penalaran seseorang ketika memberikan alasan atas pembenaran dari suatu bukti, konsep, metodologi, dan kriteria logis berdasarkan informasi atau data yang ada, dimana penalaran ini disajikan dalam bentuk argumen. *Subskillnya* meliputi menyatakan hasil, membenarkan suatu prosedur, dan menyajikan argument. Regulasi diri (*self-regulation*) merupakan secara sadar memantau aktivitas kognitif diri sendiri, unsur-unsur yang digunakan dalam kegiatan menyempurnakan hasil, terutama penerapan keterampilan dalam menganalisis dan evaluasi hasil penilaian inferensi diri sendiri dengan memandang ke arah pertanyaan, konfirmasi, validasi, atau mengoreksi salah satu alasan dari hasil seseorang. *Subskillnya* meliputi pemeriksaan dan pengkoreksian diri.

Keterampilan berpikir kritis tidak hanya dikembangkan dalam pembelajaran saja tetapi juga harus didukung dengan instrumen tes (Ennis, 1993, hlm. 180). Alat ukur yang digunakan harus sesuai dengan kompetensi yang telah ditetapkan dalam kurikulum. Banyak tipe penilaian yang dapat digunakan seperti soal pilihan ganda, pilihan ganda yang dimodifikasi, konstruksi jawaban singkat, serta uraian. Namun masih banyak pendidik yang belum menggunakan tes keterampilan berpikir kritis pada siswa.

Dalam mengembangkan suatu instrumen tes, sangatlah penting dilakukan analisis tes untuk mengetahui karakteristik tes yang dikembangkan. Karakteristik suatu tes biasanya berkaitan dengan kualitas dari tes tersebut, apakah tes yang diberikan kepada peserta didik sesuai atau tidak dengan indikator yang diharapkan. Karakteristik tes merupakan ciri khas dari tes tersebut. Jika suatu tes tidak diketahui karakteristiknya bagaimana dapat diketahui tes tersebut dapat dikatakan menjadi alat ukur yang baik atau tidak. Untuk mengetahui karakteristik dari suatu tes perlu dilakukan analisis tes, sehingga nantinya akan dapat diketahui butir tes mana yang berkualitas dan butir tes yang tidak berkualitas.

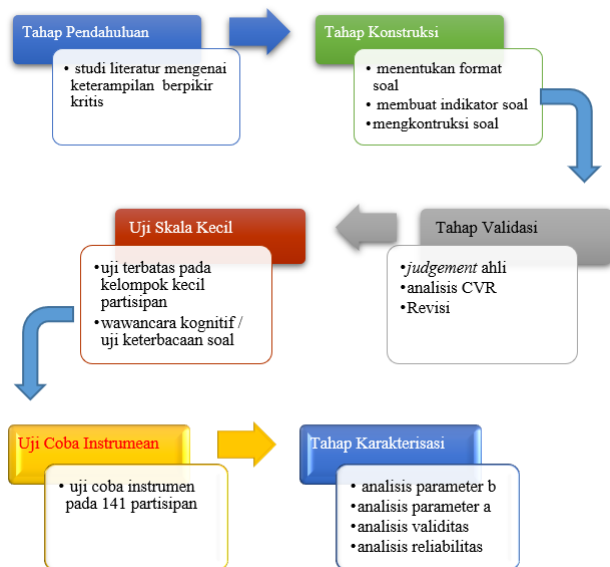
Berkaitan dengan analisis alat ukur agar mengetahui kualitas alat ukur tersebut, terdapat dua teori pengukuran yang dapat digunakan untuk analisis karakteristik tes yaitu teori tes klasik (*classical test theory* atau CTT) dan teori respon butir (*item response theory* atau IRT) (Awopeju & Afolabi, 2016, hlm. 264). Teori tes klasik berkembang sejak tahun 90-an dan masih digunakan sampai sekarang. Namun teori tes klasik memiliki keterbatasan yakni bergantung pada kelompok sampel yang digunakan akibatnya karakteristik *item* yang dianalisis dengan CTT dapat berubah sesuai konteks dari responden. Artinya suatu *item* bisa memiliki taraf kesukaran rendah karena *item* tersebut dikerjakan oleh kelompok responden dengan kemampuan tinggi namun taraf kesukaran *item* tersebut menjadi tinggi jika dikerjakan oleh kelompok responden dengan kemampuan rendah (Nurcahyo, 2016, hlm. 64). Maka dari itu dikembangkan teori baru yaitu teori respon butir atau IRT untuk mengatasi keterbatasan dari teori tes klasik.

Teori respon butir (IRT) merupakan model matematika yang memperhitungkan kemungkinan orang memberikan jawaban dengan benar untuk setiap *item* (Huang dkk., 2009, hlm. 55). Dalam teori IRT terdapat kurva utama disebut Kurva Karakteristik Butir Soal (*Item Characteristic Curve*). Kurva ini mewakili karakteristik suatu soal yang menunjukkan kemungkinan seorang peserta tes dengan kemampuan tertentu ( $\theta$ ) dapat menjawab dengan benar (Fatkhudin, dkk. 2014, hlm. 123). Dalam teori ini, model matematisnya mempunyai makna bahwa probabilitas subjek untuk menjawab butir dengan benar bergantung pada kemampuan subjek dan karakteristik butir (Retnawati, 2014, hlm. 1). Artinya, peserta tes dengan kemampuan tinggi akan memiliki probabilitas menjawab benar lebih besar dibandingkan dengan peserta tes yang memiliki kemampuan yang rendah. Maka hasil analisis dengan teori respon butir hanya dipengaruhi oleh kemampuan individu peserta tes dan butir soalnya. Karakteristik butir terdiri dari parameter daya pembeda ( $a$ ), parameter tingkat kesukaran ( $b$ ), dan faktor tebakan ( $c$ ). Ada tiga asumsi yang mendasari teori respon butir, yaitu *unidimensi*, *independensi local*, dan *invariansi* parameter (Hambleton dkk., 1991, hlm. 9-12).

Banyak peneliti yang melakukan pengembangan instrumen tes untuk mengukur keterampilan berpikir kritis. Salah satunya yaitu penelitian yang dilakukan oleh Tiruneh dkk. (2017, hlm. 669-672) mengembangkan instrumen tes keterampilan berpikir kritis yang merujuk pada pengembangan tes penilain berpikir kritis oleh Diane F Halpern pada *Halpern Critical Thinking Assesment* (HCTA) yang terdiri dari kategori memberikan alasan, menguji hipotesis, analisis argumen, kemungkinan dan analisis ketidakpastian, serta pemecahan masalah dan pembuatan keputusan. Terdiri dari 19 item soal, 7 item soal dengan format *constructed-response* dan 2 item soal dengan format *forced-choice*. Hasil akhir penelitian ini didapatkan bahwa instrumen tes berpikir kritis pada materi listrik dan magnet dapat diterima dan dapat digunakan untuk mengukur keterampilan berpikir kritis.

## METODE PENELITIAN

Penelitian ini menggunakan metode penelitian pengembangan instrumen. Partisipan dalam penelitian ini adalah 151 siswa SMP kelas IX pada empat SMP negeri dan dua SMP swasta di kota Lembang tahun pelajaran 2018/2019 yang telah mempelajari materi tekanan. Instrumen dalam penelitian ini yaitu berupa lembar *judgement* dengan aspek yang dinilai yaitu ketepatan butir soal mengukur indikator, ketepatan informasi yang diberikan dalam soal dan kejelasan kata/penyusunan kata pada butir soal serta tes keterampilan berpikir kritis yang disusun berdasarkan 16 *subskill* dari 6 aspek keterampilan berpikir kritis menurut Facione dengan format soal berupa pilihan terarah dan jawaban terstruktur.



**Gambar 1. Skema Langkah-langkah Penelitian**

Teknik pengambilan data mengikuti langkah-langkah penelitian pada gambar 1. Setelah tahap validasi, dilakukan uji coba skala kecil pada 10 partisipan untuk uji keterbacaan dan penentuan waktu pengerjaan tes. Setelah melakukan revisi maka dilakukan uji coba kepada 141 partisipan. Teknik analisis data yang digunakan yaitu analisis *judgement* ahli untuk menentukan validitas instrumen tes dan analisis hasil uji instrument tes untuk mengkarakterisasi tes berdasarkan teori respon butir dibantu menggunakan *software* IRTPRO4.2 *Student*.

Dalam suatu penelitian diperlukan sebuah instrumen tes yang valid. Validitas didefinisikan

sebagai kemampuan instrumen untuk mengukur sifat-sifat konstruk yang diteliti, hal ini merupakan faktor penting dalam pemilihan dan pengaplikasian instrumen. Validitas bukan milik dari instrumen tetapi milik skor yang diperoleh dari instrumen untuk tujuan tertentu pada sekelompok partisipan. (Zamandeh dkk., 2015, hlm. 165). Validitas isi merupakan validitas yang diestimasi lewat pengujian terhadap kelayakan atau relevansi isi tes melalui analisis rasional oleh panel yang berkompoten atau melalui expert *judgement* (penilaian ahli) (Hendryadi, 2014, hlm. 1). Jumlah ahli/pakar yang menilai relevan atau tidak dikuantisasi dengan menggunakan validitas konten rasio (CVR), formulanya ditunjukkan oleh persamaan (1) di bawah ini :

$$(N_e - \frac{N}{2}) / \frac{N}{2} \quad (1)$$

Keterangan:

$N_e$  = jumlah ahli/pakar yang menilai instrumen relevan

$N$  = jumlah ahli/pakar keseluruhan

Formula ini menghasilkan nilai-nilai yang berkisar dari -1 sampai +1, nilai positif menunjukkan bahwa setidaknya setengah panelis menilai *item* sebagai penting/esensial. Semakin lebih besar CVR dari 0, maka semakin “penting” dan semakin tinggi validitas isinya (Hendryadi, 2014, hlm. 4). Jika hasil CVR lebih besar dari 0,49 maka soal dikatakan valid atau dapat mengukur indikator dan dapat dimasukkan kedalam instrumen penelitian. Sedangkan untuk karakteristik dari hasil analisis berdasarkan teori respon butir didapatkan parameter daya pembeda (a), parameter tingkat kesukaran (b) dan reliabilitas. Untuk memahami bagaimana tes berfungsi secara keseluruhan (reliabel), dapat dilihat dari gabungan fungsi informasi tiap butir pada kurva fungsi informasi total (Edward, 2009, hlm. 522; Weinhadrt, dkk. 2012, hlm. 651).

Berdasarkan data yang diperoleh pada penelitian, maka model teori respon butir yang paling sesuai adalah menggunakan *Partial Credit Model* (PCM). Asumsi pada PCM yakni setiap butir memiliki parameter daya pembeda (a) yang sama. Bentuk umum PCM menurut

Muraki dan Book (dalam Retnawati, 2014, hlm. 37) ditunjukkan oleh persamaan (2) di bawah ini:

$$P_{jk}(\theta) = \frac{\exp[\sum_{v=0}^h(\theta_n - b_{jv})]}{\sum_{r=0}^{m_i} [\exp[\sum_{v=0}^k(\theta_n - b_{jv})]]}, k = 0, 1, 3, \dots, m \quad (2)$$

Keterangan:

$P_{jk}(\theta)$  : probabilitas peserta berkemampuan  $\theta$  memperoleh skor kategori k pada butir j

$\theta$  : kemampuan peserta

$m + 1$  : banyaknya kategori butir j

$b_{jk}$  : indeks kesukaran kategori k butir j

### HASIL DAN PEMBAHASAN

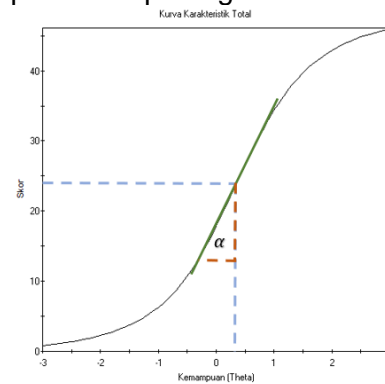
Pada tahap pendahuluan didapatkan hasil studi literatur berupa define berpikir kritis menurut beberapa ahli serta indikator atau aspek-aspek keterampilan berpikir kritisnya. Pada tahap konstruksi mendapatkan format soal yang akan dikonstruksi, indikator soal beserta soal yang telah dikonstruksi. Pada tahap validasi dilakukan judgement oleh 5 ahli. Saran perbaikan yang diperoleh dari penilai merupakan data kualitatif yang kemudian dianalisis untuk dilakukan revisi pada instrumen tes keterampilan berpikir kritis. Aspek yang dinilai dalam tahap judgement ahli yaitu relevansi soal dengan ketepatan soal mengukur indikator soal, dan kejelasan kata/penyusunan kata pada soal, serta kesesuaian informasi yang diberikan dalam soal. Hasil judgement ahli dapat dilihat pada tabel 1.

**Tabel 1.**  
Rekapitulasi hasil Validasi

No Soal	Ahl i 1	Ahl i 2	Ahl i 3	Ahl i 4	Ahl i 5	CV R	Ket	Kesimpulan
1	R	R	R	R	R	1	Valid	Revisi
2	R	R	R	R	R	1	Valid	Revisi
3	R	R	R	R	R	1	Valid	Revisi
4	R	R	R	R	R	1	Valid	Revisi
5	R	R	R	R	R	1	Valid	Revisi
6	TR	R	R	R	R	0,6	Valid	Revisi
7	TR	R	R	R	TR	0,2	Tidak Valid	Revisi
8	R	R	R	R	TR	0,6	Valid	Revisi
9	R	R	R	R	R	1	Valid	Revisi
10	TR	R	R	R	R	0,6	Valid	Tanpa revisi

No Soal	Ahl i 1	Ahl i 2	Ahl i 3	Ahl i 4	Ahl i 5	CV R	Ket	Kesimpulan
11	TR	R	R	R	R	0,6	Valid	Revisi
12	R	R	R	R	R	1	Valid	Revisi
13	TR	R	R	R	R	0,6	Valid	Tanpa revisi
14	TR	R	R	R	R	0,6	Valid	Revisi
15	R	R	R	R	R	1	Valid	Revisi
16	R	R	R	R	R	1	Valid	Tanpa revisi

Hasil uji coba skala kecil mendapatkan dua butir soal yang diperbaiki redaksi sehingga tidak membuat kebingungan partisipan dalam membaca soal. Hasil uji coba yang dilakukan kepada 141 partisipan didapatkan data yang kemudian dianalisis berdasarkan teori respon butir dengan bantuan software IRTPRO4.2 Student. Dari hasil analisis didapatkan kurva karakteristik yang menunjukkan data parameter daya pembeda (a) dan parameter tingkat kesukaran (b) serta kurva fungsi informasi total dan SEM yang menunjukkan reliabilitas tes. Kurva karakteristik total dapat dilihat pada gambar 2 serta kurva fungsi informasi total dan SEM dapat dilihat pada gambar 3.



**Gambar 2.** Kurva Karakteristik Total

a. Tingkat kesukaran

Hasil analisis parameter tingkat kesukaran (b) tiap butir dianalisis dari grafik kurva karakteristik butir. Kategori parameter tingkat kesukaran (b) dapat dilihat pada tabel 3.2. Hasil analisis menunjukkan bahwa tingkat kesukaran tiap butir berada pada rentang sedang karena berada pada rentang -1 sampai +1 Parameter tingkat kesukaran (b) tiap butir dapat dilihat pada tabel 4.9.

**Tabel 2**  
Tingkat Kesukaran Tiap Butir

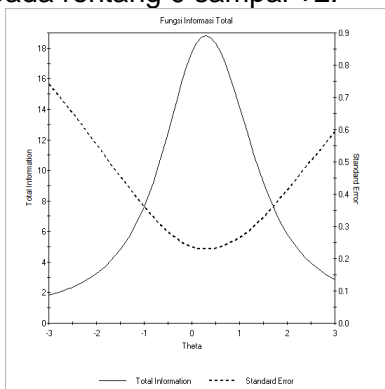
Revi Novia, Dkk- Pengembangan Dan Karakterisasi Tes Keterampilan Berpikir Kritis Materi Tekanan Berdasarkan Teori Respon Butir

Butir Soal	(b)	Butir Soal	(b)	Butir Soal	(b)	Butir Soal	(b)
1	0,92	5	0,62	9	0,69	13	0,48
2	0,34	6	0,55	10	0,88	14	0,01
3	-	7	0,25	11	0,24	15	0,22
4	0,16	8	0,03	12	0,77	16	0,10

Sedangkan perpotongan antar kategori pada tiap butir yang menunjukkan tingkat kesukaran untuk mencapai kategori berikutnya tertera pada tabel 4.10, Nilai perpotongan ini mempunyai arti bahwa peningkatan kemampuan responden untuk memberi respon kategori k ke kategori selanjutnya memiliki tingkat kesukaran  $b_{jk}$ . Hasil analisis perpotongan tiap kategori didapatkan bahwa peningkatan kemampuan partisipan untuk mencapai kategori skor selanjutnya memiliki tingkat kesukaran dari rentang sangat mudah sampai sangat sukar.

b. Daya Pembeda

Parameter daya pembeda (a) dalam PCM diasumsikan sama tiap butir karena PCM merupakan pengembangan model 1PL (Model Rush) yang hanya menganalisis parameter tingkat kesukaran saja. Berdasarkan hasil analisis kuva karakteristik total pada gambar 4.1 didapatkan bahwa parameter daya pembeda (a) tes sebesar 0,49 yang artinya tes memiliki parameter daya pembeda yang baik karena berada pada rentang 0 sampai +2.



Gambar 3. Kurva FI Total dan SEM

c. Validitas

Analisis validitas dilakukan setelah melakukan *judgement* kepada lima orang ahli. Karakteristik tes berdasarkan hasil validitas isi dapat dikatakan valid karena nilai CVR lebih dari 0.49. Salah satu butir mendapatkan nilai CVR 0,2 dikarenakan dua orang ahli memberi nilai 0 pada butir soal yang berarti bahwa indikator tidak relevan mengukur *subskill* keterampilan berpikir kritis. Tetapi salah satu ahli memberikan nilai 0 atau tidak relevan karena redaksi yang kurang tepat, sehingga butir tersebut diperbaiki dan masih tetap digunakan untuk pengambilan data. Analisis validasi dapat dilihat pada tabel 1.

Tabel 3  
Parameter tingkat kesukaran antar kategori

BS	$\delta_{i1}$	TS	$\delta_{i2}$	TS	$\delta_{i2}$	TS
1	0,4	S	0,2	S	2,1	SK
2	-0,4	S	0,9	S	0,4	S
3	-1,1	M	0,6	S	1,3	SK
4	-0,2	S	0,1	S	0	S
5	-1,0	S	0,6	S	2,3	SK
6	0,7	S	1,3	SK	-0,3	S
7	0,5	S	0,5	S	0,8	S
8	0,7	S	1,2	SK	-2,3	SM
9	0,4	S	0,8	S	0,9	S
10	0,4	S	-0,2	S	2,4	SS
11	0,3	S	0,6	S	0,2	S
12	0,8	S	1,2	SK	0,3	S
13	0,3	S	0,7	S	0,4	S
14	1,1	SK	1,0	S	-1,9	M
15	0,6	S	0,0	S	0,0	S
16	0,4	S	-0,3	S	0,2	S

Keterangan: BS: Butir Soal; TS: Tingkat Kesukaran; S: Sedang; M: Mudah; SK: Sukar; SSK: Sangat Sukar; SM; Sangat Mudah;

d. Reliabilitas

Reliabilitas tes didapatkan kurva fungsi informasi. Setiap kegiatan pengukuran pasti terdapat kesalahan penaksiran. Dalam analisis teori respon butir mengenai tes keterampilan berpikir kritis merupakan satu per-akar dari fungsi informasi yang didapat. Berdasarkan gambar 3 dapat dilihat puncak fungsi informasi total berada pada nilai 18,83 dengan kesalahan penaksiran (SEM) sebesar 0,23. Berdasarkan kurva pada gambar 3 tes keterampilan berpikir kritis yang dikembangkan terdiri dari 16 butir soal yang diberikan kepada 141 partisipan siswa menunjukkan butir-butir soal akan reliabel juga diberikan pada siswa dengan taraf



kemampuan sedang. Perpotongan antara kedua kurva juga menunjukkan bahwa butir-butir soal dalam tes keterampilan berpikir kritis akan reliabel untuk mengetahui tingkat kemampuan dalam rentang -1,10 sampai 1,7 yaitu partisipan dengan kategori kemampuan rendah hingga kemampuan tinggi.

## SIMPULAN

Berdasarkan hasil temuan dan pembahasan karakteristik soal keterampilan berpikir kritis pada materi tekanan diperoleh karakteristik tes dari 16 soal yang dikonstruksi memiliki parameter tingkat kesukaran (b) yang baik karena nilai parameter tingkat kesukaran (b) yaitu 0,34 termasuk dalam kategori sedang. Parameter daya pembeda (a) bernilai 0,49 yang berarti berada pada rentang 0 sampai +2 sehingga dikategorikan tes memiliki parameter daya pembeda (a) yang baik. Instrumen tes yang telah dikonstruksi dikatakan valid karena hasil analisis kuantitatif validasi oleh ahli menunjukkan nilai CVR lebih dari 0,49. Puncak fungsi informasi total berada pada nilai 18,83 dengan kesalahan penaksiran (SEM) sebesar 0,23 dengan perpotongan antara kedua kurva berada pada rentang -1,1 sampai 1,70 artinya instrumen tes keterampilan berpikir kritis ini valid dan reliabel untuk mengetahui tingkat keterampilan siswa dalam rentang rentang -1,10 sampai 1,70 yaitu partisipan dengan kategori keterampilan rendah hingga keterampilan tinggi.

## REFERENSI

- [1] Awopeju, O.A. & Afolibi E.R.I. (2016). Comparative Analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of Senior School Certificate Mathematics Examination. *European Scientific Journal October 2016 edition*, 12 (28). 263-284.
- [2] Bybee, R.W. & Fuchs, B. (2006). Preparing the 21st Century Workforce: A New Reform in Science and Technology Education. *Journal Of Research In Science Teaching*, 43. 349-352.
- [3] Edwards, M.C. (2009). An Introduction to Item Response Theory Using the Need for Cognition Scale. *Social and Personality Psychology Compass*, 3 (4). 507-529.
- [4] Ennis, R.H. (1993). *Critical Thinking Assessment*. USA: The Ohio State University.
- [5] Facione, P.A. (2013). *Critical Thinking: What It Is and Why It Counts*. California : The California Academic Press.
- [6] Fatkhudin, A., Surarso, B., Subagio, A. (2014). Item Response Theory Model Empat Parameter Logistik Pada Computerized Adaptive Test. *Jurnal Sistem Informasi Bisnis*, 02. 121-129.
- [7] Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of Item Response Theory*. Newbury Park, CA: Sage Publication Inc.
- [8] Hendryadi. (2014). Content Validity (Validitas Isi). *Teorioline Personal Paper*, (01). 1-5.
- [9] Huang, Y.M., Lin, Y.T., & Cheng, S.C. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers & Education* 52. 54-67.
- [10] Nurcahyo, F.A. (2016). Aplikasi IRT dalam Analisis Aitem Tes Kognitif. *Buletin Psikologi*, 24 (2). 64-75.
- [11] Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya*. Yogyakarta: Nuha Medika.
- [12] Rofiah, E., Aminah, N.S., Ekawati, E.Y. Penyusunan Instrumen Tes Kemampuan Berpikir Tingkat Tinggi Fisika Pada Siswa SMP. *Jurnal Pendidikan fisika*, 1 (2). 17-22.
- [13] Tiruneh, D.T., Cock, M.D., 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). 663-682.
- [14] Tiruneh, D.T., Vanburgh, E., Elen, J. (2014). Effectiveness of Critical Thinking Instruction in Higher Education: A Systematic Review of Intervention Studies. *Higher Education Studies*, 4 (1). 1-17.
- [15] Weinhardt, J.M., Morse, B.J., Chimeli, J., Fische, J. (2012). An item response theory and factor analytic examination of two

prominent maximizing tendency scales. *Judgement and Decision Making*, 7 (5). 644-658.

- [16] Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Jurnal of Caring Sciences*, 4 (2). 165-178.