

# Karakterisasi Tes Penalaran Ilmiah Materi Suhu Dan Kalor Berdasarkan Teori Respon Butir

Wilianti Saptawulan, Taufik Ramlan Ramalis, dan Muslim\*

Program Studi Pendidikan Fisika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi No.229, Bandung 40154, Indonesia

\*Corresponding author's e-mail: muslim@upi.edu

## ABSTRAK

Penalaran ilmiah merupakan salah satu keterampilan yang penting dimiliki oleh setiap siswa. Penelitian ini bertujuan untuk menghasilkan produk berupa instrumen tes penalaran ilmiah materi suhu dan kalor dan untuk mengetahui karakteristik instrumen tes berdasarkan teori respon butir. Metode yang digunakan dalam penelitian ini adalah metode kombinasi desain *concurrent embedded*. Data kualitatif diperoleh dari *judgement* ahli dan data kuantitatif diperoleh dari hasil uji instrumen tes yang dibuat oleh peneliti dengan mengacu pada *lawson's classroom test of scientific reasoning* (LCTSR). Partisipan penelitian berjumlah 139 siswa yang dipilih secara acak dari 3 sekolah menengah pertama di Kota Cilegon dengan kriteria pernah mempelajari materi suhu dan kalor. Hasil penelitian menunjukkan bahwa instrumen tes yang dibuat memiliki daya pembeda dalam kategori baik, tingkat kesukaran dalam kategori sedang, dan faktor tebakan semu dalam kategori baik. Tes penalaran ilmiah ini memiliki fungsi informasi sebesar 3,695 dengan sem 0,520. Dapat disimpulkan bahwa tes penalaran ilmiah sesuai jika diberikan kepada siswa dengan kemampuan sedang dan tes penalaran ilmiah ini akan reliabel jika diberikan pada siswa dengan kemampuan rendah hingga kemampuan tinggi.

**Kata kunci:** Karakterisasi Tes, Penalaran Ilmiah, Teori Respon Butir

## 1. Pendahuluan

Penalaran adalah proses menarik kesimpulan dari prinsip-prinsip dan bukti untuk membuat kesimpulan baru atau mengevaluasi kesimpulan yang diajukan [1]. Dari perspektif literasi sains Bao (2006) menyatakan bahwa penalaran ilmiah merupakan keterampilan kognitif yang diperlukan untuk memahami dan mengevaluasi informasi ilmiah, yang sering melibatkan pemahaman dan mengevaluasi teoritis, hipotesis statistik, dan kausal [2]. Shermer (2002) mengungkapkan bahwa penalaran ilmiah adalah seperangkat metode yang dirancang untuk menggambarkan dan menginterpretasikan pengamatan atau menyimpulkan fenomena, masa lalu atau sekarang, dan bertujuan menguji bidang pengetahuan sebagai penolakan atau konfirmasi [3].

Dalam sudut pandang yang lebih operasional, penalaran ilmiah dinilai berkaitan dengan serangkaian keterampilan penalaran dasar yang biasanya dibutuhkan agar siswa berhasil melakukan penyelidikan ilmiah,

termasuk menjajaki masalah, merumuskan dan menguji hipotesis, memanipulasi dan mengisolasi variabel, dan mengamati dan mengevaluasi konsekuensinya [4].

Han (2013) menyatakan bahwa keterampilan penalaran ilmiah penting karena berkaitan dengan domain Pendidikan [4]. Keterampilan penalaran ilmiah adalah alat yang memungkinkan seseorang untuk memperoleh pengetahuan baru dan berpikir kritis. Selain itu Bao (2009) mengemukakan bahwa keterampilan penalaran ilmiah memiliki hubungan dengan keterampilan berpikir kritis dan kemampuan bernalar yang sesuai dengan aspek pengetahuan pada *science, technology, engineering, and mathematics* (STEM) serta sangat penting dimiliki oleh setiap siswa [2].

Terdapat beberapa jenis format tes yang digunakan untuk mengukur penalaran ilmiah siswa, salah satunya adalah format tes yang dikembangkan oleh Lawson (1978) yang memiliki sebutan LTR (Lawson Test of Scientific Reasoning) yang sekarang disebut sebagai CTR (Classroom Test of Scientific

*Reasoning*) (2000). Lawson (2000) mengembangkan instrumen yang terdiri dari beberapa dimensi penalaran diantaranya penalaran proporsional, penalaran korelasional, penalaran kontrol variabel, dan penalaran hipotesis-deduktif. Dalam penelitian ini instrumen tes penalaran yang digunakan merupakan seperangkat soal pilihan ganda berjumlah 13 butir yang mengacu pada empat dimensi penalaran tes Lawson yang telah *dijudgement* ahli.

Penelitian terdahulu pernah dilakukan oleh Colleta dan Philips [5] mengenai konsep gaya dengan mengembangkan tes penalaran milik Lawson. Penelitian lainnya juga dilakukan oleh Stephens dan Clement [6] pada materi gerak dan gaya. Padahal banyak materi fisika lain yang dapat dijadikan pengukuran penalaran ilmiah seseorang, seperti materi suhu dan kalor. Materi suhu dan kalor dipilih dalam penelitian ini karena dianggap sebagai salah satu materi dasar dalam fisika dan kontekstual dalam kehidupan sehari-hari. Selain itu, berdasarkan studi literatur belum ada yang mengembangkan instrumen tes penalaran ilmiah dengan topik materi suhu dan kalor.

Ketika sebuah tes telah diujikan pada sampel, maka diperlukan suatu analisis tes untuk mengetahui hasil dari uji tes. Hasil dari uji tes akan mengkarakterisasi tes yang diujikan. Terdapat dua jenis analisis tes yang masih digunakan sampai saat ini yakni analisis tes menggunakan teori tes klasik dan analisis tes dengan menggunakan teori tes modern (*item response theory/IRT*). Keunggulan analisis tes secara klasik terletak pada kemudahan dalam penggunaannya. Namun kelemahan teori ini memiliki ciri bahwa karakteristik butir tes tidak dapat dipisahkan dari karakteristik peserta tes karena hasil estimasi parameter tergantung pada karakteristik peserta tes (*group dependent*) dan

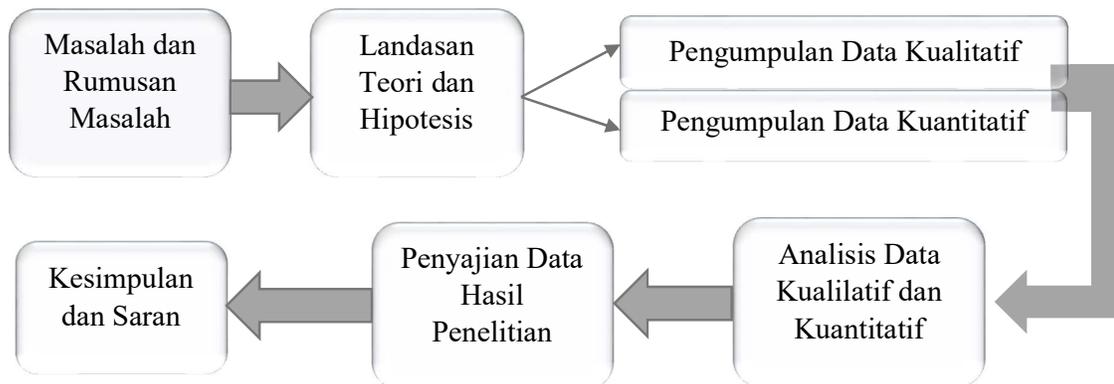
hasil estimasi kemampuan peserta yang bergantung pada karakteristik butir.

Adapun kelebihan IRT adalah bahwa: (1) IRT tidak bergantung grup, (2) skor peserta didik dideskripsikan bukan *test dependent*, (3) model ini menekankan pada tingkat butir soal bukan tes, (4) IRT tidak memerlukan paralel tes untuk menentukan reliabilitas tes, (5) IRT suatu model yang memerlukan suatu pengukuran ketepatan untuk setiap skor tingkat kemampuan [7]. Selain itu dalam penggunaannya juga analisis IRT dapat digunakan pada menu *add-ins Microsoft Excel*.

Terdapat tiga asumsi teori respon butir, yaitu 1) unidimensi, setiap butir tes hanya mengukur satu kemampuan, 2) independensi lokal, tidak ada korelasi antara peserta tes dengan peserta tes lainnya pada butir soal yang sama maupun berbeda, dan 3) invariansi parameter, karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal [7].

Berdasarkan respon instrumen teori respon butir terbagi menjadi dua jenis model respon, yakni model dikotomi dan model politomi. Model dikotomi berlaku ketika pilihan jawaban instrumen hanya terdapat dua pilihan jawaban, seperti iya atau tidak, benar atau salah, maupun setuju atau tidak setuju. Jawaban pada model dikotomi memiliki skor 1 dan 0. Sedangkan untuk model politomi memiliki respon instrumen yang bervariasi seperti pada kuesioner yang diberikan skala 1-5 ataupun dengan pilihan jawaban dari sangat tidak setuju, tidak setuju, setuju hingga sangat setuju.

Retnawati (2014) memaparkan bahwa ada 3 model yang dapat digunakan untuk melakukan analisis dengan menggunakan teori respon butir model dikotomi, yakni model 1



Gambar 1. Desain penelitian *Concurrent Embedded*

parameter logistik, 2 parameter logistik, dan 3 parameter logistik (yang kemudian akan digunakan istilah: 1 PL, 2 PL, dan 3 PL) [8]. Pengguna teori ini perlu memilih, data yang dianalisis apakah sesuai dengan salah satu dari ketiga model tersebut berdasarkan fungsi informasi.

Model 1 parameter logistik hanya menitikberatkan pada parameter tingkat kesukaran butir soal (parameter b). Model 1 PL ini juga disebut sebagai model Rasch [7]. Model 2 parameter logistik melibatkan dua parameter diantaranya tingkat kesukaran (b) dan daya pembeda butir (a). Model 3 parameter logistik melibatkan tiga parameter yakni tingkat kesukaran (b), daya pembeda butir (a), dan faktor tebakan (*guessing*) (c).

Untuk mengetahui kekuatan suatu tes maka diperlukan reliabilitas tes. Reliabilitas merupakan derajat keajegan hasil pengukuran pada objek yang sama, jika dilakukan pengukuran mengenai kemampuan seorang siswa, hasil pengukurannya akan sama meskipun pengujian dan butir soal berbeda namun memiliki karakteristik yang sama [8]. Estimasi reliabilitas tes berdasarkan teori respon butir diperoleh dengan fungsi informasi dan kesalahan pengukurannya atau *standard error measurement* (SEM). Fungsi informasi memiliki hubungan yang berlawanan dengan SEM, semakin rendah fungsi informasi sebuah butir, maka semakin besar SEM pada butir tersebut. Semakin kecil SEM pengukuran akan semakin tepat, reliabel, dan dapat dipercaya hasil pengukurannya [9].

## 2. Metode Penelitian

Metode penelitian yang digunakan yaitu metode kombinasi dengan desain *Concurrent Embedded*. Metode kombinasi dengan desain *Concurrent Embedded* adalah metode penelitian yang menggabungkan antara metode kuantitatif dan metode kualitatif dengan cara mencampur metode tersebut secara tidak seimbang [10].

Penelitian ini dilakukan dengan 7 tahap seperti yang disajikan pada Gambar 1. Tahap pertama dengan melakukan studi pendahuluan untuk mengetahui masalah dan untuk merumuskan masalah. Tahap kedua dengan studi literatur untuk mengetahui teori yang berkaitan dan untuk membuat hipotesis. Setelah tahap kedua peneliti mengembangkan instrumen tes berdasarkan teori yang didapat dari studi literatur. Tahap yang ketiga adalah mengumpulkan data kualitatif. Pengumpulan data kualitatif dilakukan dengan cara melakukan

judgement ahli terhadap instrumen yang telah dibuat. Hasil judgement dapat berupa catatan saran dan skor. Catatan saran penilai akan menjadi data kualitatif dan skor judgement akan menjadi data kuantitatif. data kuantitatif juga didapat dari hasil uji lapangan instrumen tes. Kemudian tahap keempat yaitu menganalisis data. Hasil validasi dari beberapa ahli akan diolah dengan menerapkan metode validasi Aiken'V yang nantinya dapat diketahui bahwa tes tersebut valid atau tidak valid. Sedangkan hasil yang didapat dari pengumpulan data kuantitatif kemudian dianalisis berdasarkan Teori Respon Butir dengan menggunakan aplikasi eirt 2.0 yang terpasang pada menu add-ins Microsoft Excel. Tahap keenam yaitu penyajian data hasil analisis serta pembahasannya. Terakhir dengan menyajikan simpulan dan saran dari penelitian.

Partisipan penelitian dipilih secara *purposive sampling* dalam hal ini adalah peserta didik pada tingkat sekolah menengah pertama yang sudah mempelajari materi Suhu dan Kalor dan pemilihan kelas didasarkan atas rekomendasi dari guru fisika di sekolah partisipan. Partisipan dalam penelitian ini yaitu 48 siswa dari SMPN 1 Cilegon, 32 siswa dari SMPN 2 Cilegon, dan 59 siswa dari SMP IT Raudhatul Jannah Cilegon. Sehingga jumlah total partisipan dalam penelitian ini yaitu 139 orang.

## 3. Hasil Dan Pembahasan

Pada tahap studi pendahuluan dilakukan studi literatur untuk menemukan dan merumuskan masalah. Studi literatur juga dilakukan untuk mengetahui instrumen yang sesuai dan dapat digunakan untuk mengukur penalaran ilmiah serta untuk mengetahui hasil penelitian-penelitian yang telah dilakukan selama ini mengenai tes penalaran ilmiah sebagai perbandingan. Selain itu dari studi literatur juga dapat diketahui cara mengukur validasi tes yang telah dibuat. Adapun untuk mengetahui analisis tes berdasarkan teori respon butir menggunakan software eirt yang sudah terpasang pada *toolbar add-ins Microsoft excel*. Hasil dari studi literatur menjadi landasan teori dalam penelitian.

Tahap selanjutnya adalah pembuatan instrumen tes. Instrumen tes yang dibuat berupa seperangkat tes penalaran ilmiah materi suhu dan kalor dengan mengacu pada 4 dimensi penalaran Lawson. Dimensi penalaran yang digunakan yaitu penalaran proporsional,

penalaran korelasional, penalaran kontrol variabel, dan penalaran hipotesis-deduktif.

### 3.1 Validitas

Validasi ahli dianalisis berdasarkan metode Aiken'V untuk mengetahui kesepakatan hasil validasi. Berdasarkan validasi juga dapat diketahui butir mana yang harus diperbaiki sehingga butir valid dan dapat diujikan. Hasil *judgement* para ahli dalam penelitian ini diolah berdasarkan metode Aiken'V dengan menggunakan persamaan [11]:

$$V = \Sigma S / [n(c - 1)] \quad (1)$$

$$S = r - lo \quad (2)$$

Keterangan

V : koefisien validitas isi

lo : angka penilaian validitas yang terendah (misalnya 1)

c : angka penilaian validitas tertinggi (misalnya 4)

r : angka yang diberikan oleh penilai

Koefisien validitas isi (V) memiliki kemungkinan 0 sampai dengan 1, nilai tersebut menunjukkan derajat dari validitas butir. Sebuah butir dapat dianggap valid jika nilai  $V \geq 0,5$  [12]. Interpretasi indeks validasi ahli dapat juga menggunakan kriteria sesuai dengan Tabel 1.

Tabel 1. Interpretasi Aiken'V

Hasil Validasi	Kriteria
$0,80 < V \leq 1,00$	Sangat Tinggi
$0,60 < V \leq 0,80$	Tinggi
$0,40 < V \leq 0,60$	Cukup
$0,20 < V \leq 0,40$	Rendah
$0,00 < V \leq 0,20$	Sangat Rendah

Untuk interpretasi lebih rinci dari hasil *judgement* ahli disajikan pada Tabel 2. dengan A, B, dan C merupakan penilai (*judge*). Berdasarkan hasil *judgement* ahli dengan menggunakan metode Aiken'V dapat diketahui terdapat 11 butir soal dengan tingkat kevalidan sangat tinggi dan 2 butir soal dengan tingkat kevalidan tinggi. Selain itu dari analisis secara kualitatif terdapat beberapa soal harus diperbaiki karena terdapat kesalahan penulisan, pemilihan kata yang kurang sesuai, ilustrasi gambar yang kurang jelas, dan pilihan jawaban yang harus disesuaikan dengan pertanyaan. Sehingga beberapa butir dilakukan perbaikan.

### 3.2 Hasil Uji Lapangan

Hasil uji lapangan dianalisis berdasarkan teori respon butir (item response theory/IRT). Model yang digunakan untuk menganalisis adalah model dikotomi 3 parameter logistik.

Hasil uji lapangan digunakan untuk mengetahui estimasi karakteristik tes dan estimasi kemampuan partisipan.

Hasil estimasi karakteristik tes disajikan sesuai dengan model logistik 3 parameter butir yaitu daya pembeda (a), tingkat kesukaran (b), dan tebakan semu (c). Pada kurva karakteristik, parameter a ditunjukkan oleh kemiringan kurva. Semakin besar kemiringan kurva maka semakin besar daya pembeda butir [8]. Secara teoretis, nilai a terletak antara  $-\infty$  dan  $+\infty$ . Pada butir yang baik nilai ini mempunyai hubungan positif dengan performen pada butir dan a terletak antara 0 dan 2 [13].

Parameter b merupakan suatu titik pada skala kemampuan agar peluang menjawab benar sebesar 50%. Semakin besar nilai parameter b, maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50%. Dengan kata lain, semakin besar nilai parameter b maka semakin sulit butir soal tersebut [14]. Suatu butir dikatakan baik jika nilai b berkisar antara -2 dan +2, jika nilai b mendekati -2 maka indeks kesukaran butir sangat rendah sedangkan jika nilai b mendekati +2 maka indeks kesukaran butir sangat tinggi [13]. Dan nilai c termasuk dalam kategori baik jika  $c < 1/k$ , dengan k adalah banyaknya pilihan jawaban.

Secara matematis model logistik tiga parameter dapat dituliskan sebagai berikut:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \quad (3)$$

Keterangan

$P_i(\theta)$  : probabilitas peserta tes yang memiliki kemampuan  $\theta$  dipilih secara acak dapat menjawab butir I dengan benar

$\theta$  : tingkat kemampuan subjek (sebagai variabel bebas)

$a_i$  : indeks daya pembeda butir ke-i

$b_i$  : indeks kesukaran butir ke-i

$c_i$  : indeks tebakan semu butir ke-i

e : bilangan natural yang nilainya mendekati 2,718

n : banyaknya butir dalam tes

Berdasarkan hasil analisis dengan teori respon butir didapat hasil estimasi parameter butir model 3 PL sebagai berikut:

Tabel 2.  
Hasil Analisis Judgement Ahli berdasarkan Aiken'V

Penilai Butir	A	B	C	$\Sigma s$	n(c-1)	V	Interpretasi
1	Skor = 3; s = 2	Skor = 4; s = 3	Skor = 4; s = 3	8	9	0,89	Sangat Tinggi
2	Skor = 2; s = 1	Skor = 4; s = 3	Skor = 4; s = 3	7	9	0,78	Tinggi
3	Skor = 3; s = 2	Skor = 4; s = 3	Skor = 4; s = 3	8	9	0,89	Sangat Tinggi
4	Skor = 4; s = 3	Skor = 4; s = 3	Skor = 4; s = 3	9	9	1	Sangat Tinggi
5	Skor = 3; s = 2	Skor = 4; s = 3	Skor = 4; s = 3	8	9	0,89	Sangat Tinggi
6	Skor = 4; s = 3	Skor = 4; s = 3	Skor = 4; s = 3	9	9	1	Sangat Tinggi
7	Skor = 4; s = 3	Skor = 4; s = 3	Skor = 4; s = 3	9	9	1	Sangat Tinggi
8	Skor = 4; s = 3	Skor = 4; s = 3	Skor = 4; s = 3	9	9	1	Sangat Tinggi
9	Skor = 4; s = 3	Skor = 4; s = 3	Skor = 4; s = 3	9	9	1	Sangat Tinggi
10	Skor = 3; s = 2	Skor = 4; s = 3	Skor = 3; s = 2	7	9	0,78	Tinggi
11	Skor = 3; s = 2	Skor = 4; s = 3	Skor = 4; s = 3	8	9	0,89	Sangat Tinggi
12	Skor = 3; s = 2	Skor = 4; s = 3	Skor = 4; s = 3	8	9	0,89	Sangat Tinggi
13	Skor = 3; s = 2	Skor = 4; s = 3	Skor = 4; s = 3	8	9	0,89	Sangat Tinggi

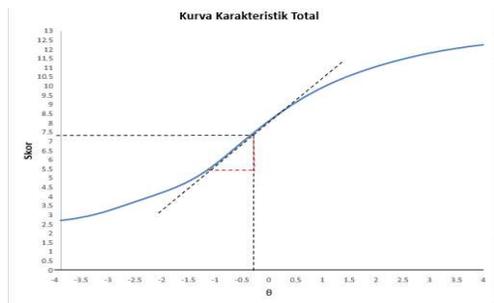
Tabel 3.  
Hasil estimasi parameter butir model 3 PL

Butir	Daya Pembeda (a)	Kriteria	Tingkat Kesukaran (b)	Kriteria	Tebakan semu (c)	Kriteria
1	1.125	Baik	0.426	Baik	0.171	Baik
2	0.438	Baik	-2.474	Tidak Baik	0.170	Baik
3	1.202	Baik	2.367	Tidak Baik	0.143	Baik
4	0.330	Baik	3.246	Tidak Baik	0.195	Baik
5	2.707	Tidak Baik	-2.854	Tidak Baik	0.166	Baik
6	1.610	Baik	-1.263	Baik	0.160	Baik
7	3.250	Tidak Baik	-0.844	Baik	0.149	Baik
8	1.348	Baik	0.305	Baik	0.187	Baik
9	1.702	Baik	0.000	Baik	0.200	Baik
10	1.304	Baik	-0.547	Baik	0.157	Baik
11	0.491	Baik	1.193	Baik	0.186	Baik
12	1.913	Baik	-0.105	Baik	0.124	Baik
13	0.773	Baik	0.842	Baik	0.167	Baik

Berdasarkan Tabel 3 dapat diketahui bahwa untuk daya pembeda ( $a$ ) terdapat 11 butir termasuk dalam kategori baik dan 2 butir tidak baik. Untuk tingkat kesukaran terdapat 9 butir dalam kategori baik dan 4 butir tidak baik termasuk sangat mudah atau sangat sukar. Sedangkan untuk tebakan semu keseluruhan butir termasuk dalam kategori baik.

Dalam analisis teori respon butir hasil analisis karakterisasi tes secara keseluruhan dapat dilihat dari kurva karakteristik total (TCC). Kurva karakteristik total akan menunjukkan hasil estimasi skor yang diperoleh partisipan dari mengerjakan 13 butir soal untuk setiap kemampuan. Kurva karakteristik total model 3 PL disajikan pada Gambar 2.

Partisipan dengan tingkat kemampuan -4 (sangat rendah) akan memperoleh skor 3 dari skor total 13, artinya hanya mampu menjawab benar 3 butir soal dari jumlah total 13 butir soal. Sedangkan partisipan dengan kemampuan 4 (sangat tinggi) akan memperoleh skor 12 dari skor total 13, artinya partisipan dengan kemampuan sangat tinggi akan mampu menjawab benar 12 butir soal dari jumlah total 13 butir soal. Maka dapat disimpulkan bahwa kemampuan partisipan dari rentang -4 sampai dengan 4, rentang skor yang diperoleh yaitu 3 sampai 12.



Gambar 2. Kurva karakteristik total

Probabilitas 1 berada pada skor 12 dan probabilitas 0 berada pada skor 3, artinya probabilitas 0,5 berada di skor 7,5 pada kurva karakteristik total. Berdasarkan gambar 4.36 dapat diketahui nilai parameter  $b$  dengan cara menarik garis horizontal pada nilai probabilitas 0,5 sampai kurva karakteristik total. Kemudian garis vertikal dibuat tepat pada perpotongannya. Nilai yang ditunjukkan pada sumbu  $X$  merupakan nilai parameter  $b$  dari TCC. Nilai parameter  $b$  atau tingkat kesukaran tes penalaran ilmiah berdasarkan kurva

karakteristik total yaitu sebesar -0,25 atau dalam kategori tingkat kesukaran sedang. Nilai parameter  $c$  merupakan asimtot dari kurva karakteristik yang mempresentasikan asumsi faktor tebakan semu dari partisipan. Berdasarkan kurva karakteristik total diketahui bahwa nilai parameter  $c$  untuk tes penalaran ilmiah berada pada skor 3 sehingga probabilitasnya adalah 0,231 artinya berada dalam kategori baik karena nilai parameter  $c$  berada pada nilai kurang dari  $1/k$  yaitu kurang dari 0,25. Sedangkan nilai parameter  $a$  diperoleh dari kemiringan lereng kurva atau dapat diketahui dengan  $\tan \alpha$ . Berdasarkan kurva karakteristik total diperoleh kemiringan kurva sebesar  $45^\circ$ , sehingga  $\tan \alpha = \tan 45^\circ = 1,00$  artinya daya pembeda (parameter  $a$ ) tes penalaran ilmiah sebesar 1,00 dalam kategori baik.

### 3.3 Fungsi Informasi dan SEM

Reliabilitas merupakan derajat keajegan hasil pengukuran pada objek yang sama, jika dilakukan pengukuran mengenai kemampuan seorang siswa, hasil pengukurannya akan sama meskipun pengujian dan butir soal berbeda namun memiliki karakteristik yang sama [8]. Estimasi reliabilitas tes berdasarkan teori respon butir diperoleh dengan fungsi informasi dan kesalahan pengukurannya atau standard error measurement (SEM). Secara matematis, fungsi informasi butir memenuhi persamaan sebagai berikut.

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (4)$$

Keterangan

$i$  : 1, 2, 3, ...,  $n$

$I_i(\theta)$  : fungsi informasi butir ke- $i$

$P_i(\theta)$  : peluang peserta dengan kemampuan  $\theta$  menjawab benar butir  $i$

$P'_i(\theta)$  : turunan fungsi  $P_i(\theta)$  terhadap  $\theta$

$Q_i(\theta)$  : peluang peserta dengan kemampuan  $\theta$  menjawab benar butir  $i$

Fungsi informasi tes merupakan jumlah dari fungsi informasi butir penyusun tes tersebut [13]. Berhubungan dengan hal ini, fungsi informasi perangkat tes akan tingkat tinggi jika butir tes mempunyai fungsi informasi yang tinggi pula. Fungsi informasi perangkat tes secara matematis dapat dituliskan sebagai berikut.

$$FI_i(\theta) = \sum_{i=1}^n I_i(\theta) \quad (5)$$

Keterangan

$I$  : 1, 2, 3, ...,  $n$

$I_i(\theta)$  : fungsi informasi butir ke-i

$FI_i(\theta)$  : fungsi informasi tes

Dalam teori respon butir, kesalahan penaksiran standar (*Standard Error of Measurement*, SEM) berkaitan erat dengan fungsi informasi. Fungsi informasi dengan SEM mempunyai hubungan yang berbanding terbalik kuadrat, semakin besar fungsi informasi maka semakin kecil atau sebaliknya [7]. Secara matematis, nilai estimasi SEM dapat dinyatakan dengan persamaan berikut.

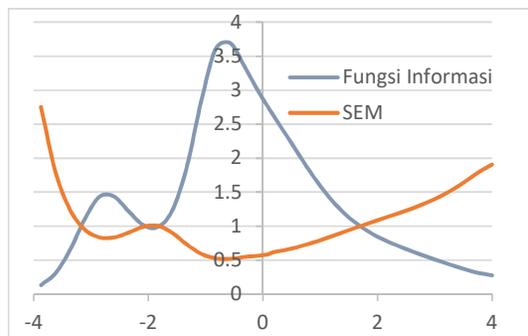
$$SEM(\hat{\theta}) = \frac{1}{\sqrt{FI(\hat{\theta})}} \quad (6)$$

Keterangan

$SEM(\hat{\theta})$  : kesalahan baku dalam pengukuran

$FI_i(\theta)$  : fungsi informasi tes

Hasil pengolahan data hubungan antara fungsi informasi dengan SEM pada model 3 PL disajikan oleh Gambar 3



Gambar 3. Fungsi informasi dan SEM model 3PL

Puncak fungsi informasi total model 3 PL berada pada nilai 3,695 dengan kesalahan penaksiran (SEM) sebesar 0,520. Berdasarkan kurva pada Gambar 3 tes penalaran ilmiah yang dikembangkan terdiri dari 13 butir soal yang diberikan kepada 139 partisipan siswa menunjukkan butir-butir soal akan reliabel jika diberikan pada siswa dengan taraf kemampuan sedang. Perpotongan antara kedua kurva juga menunjukkan bahwa butir-butir soal dalam tes penalaran ilmiah akan cocok untuk mengetahui tingkat kemampuan partisipan dalam rentang -1,84 sampai 1,71 yaitu partisipan dengan kategori kemampuan rendah hingga kemampuan tinggi.

#### 4. Simpulan

Simpulan yang dapat diambil berdasarkan hasil penelitian yang dilakukan adalah tes keterampilan penalaran ilmiah materi suhu dan kalor memiliki karakteristik tes yang baik dan akan reliabel jika diberikan pada siswa dengan tingkat kemampuan sedang.

#### 5. Referensi

- [1] Lee, C.-Q., & She, H.-C. (2010). Facilitating Students' Conceptual Change and Scientific Reasoning Involving the Unit of Combustion. *Research Science Education*, 40, 479-504
- [2] Bao, L. Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L., Luo, Y., Wang, Y., Li, L., & Wu, N. (2009). Learning and Scientific Reasoning. *Science AAAS*, 323 (586) 1-9.
- [3] Shermer, M. (2002). *Why People Believe Weird Things: Pseudo -Science, Superstition, And Bogus Notions Of Our Time*. New York, NY: Henry Holt and Company, LLC.
- [4] Han, J. (2013). *Scientific Reasoning: Research, Development, and Assessment*. (Disertasi). Ohio State: The Ohio State University
- [5] Coletta, P.V. & Phillips, A.J. (2005) Interpreting FCI scores: Normalized gain, pre-instruction scores, and scientific reasoning ability. *American Journal of Physics*.
- [6] Stephens, A.L. & Clement, J.J. (2010). Documenting the use of expert scientific reasoning processes by high school physics students. *Physical Review Special Topics-Physics Education Research* 6, 020122. DOI: 10.1103/PhysRevSTPER.6.020122
- [7] Hambleton, R.K., Swaminathan, H & Rogers, H.J. (1991). *Fundamental of Item Response Theory*. Newbury Park, CA: Sage Publication Inc.
- [8] Retnawati, H. (2014). *Teori Respon Butir dan Penerapannya*. Yogyakarta: Nuha Medika.
- [9] Setiawati, F. A., Mardapi, D., & Azwar, S., (2013). Penskalaan Teori Klasik Instrumen Multiple Intelligences Tipe Thurstone dan Likert. *Jurnal Penelitian dan Evaluasi Pendidikan* vol 17, No 2.
- [10] Sugiyono, (2016). *Metode Penelitian Kombinasi*. Bandung: Alfabeta

- [11] Hendryadi. (2014). Content Validity (Validitas Isi). *Teorioline Personal Paper*. No. 01
- [12] Suseno, M. N. (2014). Pengembangan Pengujian Validitas Isi dan Validitas Konstrak: Interpretasi Hasil Pengujian Validitas. *Proceeding Seminar Nasional Psikometri*. 70-83. Yogyakarta: Publikasi Ilmiah
- [13] Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer
- [14] Sainuddin, S. (2018). Analisis Karakteristik Butir Tes Matematika Pada Tes Buatan MGMP Matematika Kota Makassar Berdasarkan Teori Modern (Teori Respon Butir. *Jurnal Penelitian Matematika dan Pendidikan Matematika*. Vol 1 No 1.