



Kajian implementasi teori respon butir dalam menganalisis instrumen tes materi fisika

Asri Andayani*, Purwanto, Taufik Ramlan Ramalis

Program Studi Pendidikan Fisika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia.

*e-mail: asriandayani@student.upi.edu

Abstrak

Hasil studi pendahuluan yang dilakukan melalui wawancara terhadap beberapa guru fisika di kota Bandung menunjukkan bahwa kebanyakan guru tidak melakukan analisis terhadap tes yang diberikan kepada murid dan sedikit guru yang melakukan analisis serta masih menggunakan teori analisis klasik. Teori respon butir merupakan teori analisis butir soal yang termasuk ke dalam teori tes modern dan merupakan perbaikan dari teori klasik. Berdasarkan kajian literatur, teori respon butir memiliki keunggulan daripada teori klasik yaitu diantaranya dapat menggambarkan lebih rinci terkait karakteristik butir soal serta tidak memiliki ketergantungan hasil analisis pada responden. Namun teori respon butir memiliki perhitungan matematis yang cukup rumit sehingga dalam penggunaannya memerlukan bantuan *software* khusus. Metode penelitian yang dilakukan yaitu metode deskriptif dengan kajian literatur yang bertujuan untuk mengetahui efektivitas penggunaan teori respon butir dalam menganalisis instrumen dengan menggunakan *software* yang terhubung pada *Microsoft Excel* seperti *IRTpro*, dan sebagainya. Beberapa makalah ilmiah yang diterbitkan pada jurnal menunjukkan bahwa pada teori respon butir untuk menentukan karakteristik butir soal digambarkan melalui ICC (*Item Characteristics Curve*), serta akan didapat fungsi informasi dan SEM (*Standard Error Measurement*) untuk melihat reliabilitas dan validitas tes yang diukur.

Kata kunci: Analisis Instrumen; Teori Analisis; Teori Respon Butir

1. Pendahuluan

Evaluasi merupakan tahapan penting dalam pembelajaran. Secara filosofis, proses evaluasi merupakan proses penting dalam tiga mata jangkar pendidikan yang terdiri dari tujuan, metode dan evaluasi (Sutiadi, A., & Kurniawati, R. 2015). Salah satu komponen wajib dalam evaluasi yaitu instrumen tes. Setelah melewati tahap validasi isi, selanjutnya instrumen akan melalui tahap validasi empirik dengan mengujicobakan pada sejumlah siswa. Instrumen tes yang baik dan sesuai dapat diketahui melalui karakteristik dari tes itu sendiri dengan cara melakukan analisis baik terhadap butir maupun tes secara keseluruhan, sehingga akan diketahui butir tes yang baik dan butir tes yang tidak baik. Analisis butir soal atau analisis soal (*item analysis*) merupakan suatu prosedur sistematis, yang akan memberikan informasi-informasi yang sangat khusus terhadap butir soal yang telah dibuat (Arikunto, S., 2012).

Hasil studi lapangan yang dilakukan dengan cara mewawancarai guru mata

pelajaran fisika di Kota Bandung didapati bahwa permasalahan umum yang terjadi di sekolah adalah bahwa instrumen tes yang diberikan kepada siswa kebanyakan tidak melalui tahap analisis tes untuk mengetahui karakteristik dari tes tersebut dan sebagian kecil yang melakukan analisis serta masih menggunakan teori analisis klasik atau disebut *classical test theory* (CTT).

Teori analisis klasik kurang mampu merefleksikan kemampuan siswa (sebagai peserta tes) yang sebenarnya (Amelia, R.N. & Kriswanto, 2017). Maka dari itu, model teori tes modern atau disebut *Items Response Theory* (IRT) hadir untuk melengkapi keterbatasan dalam pengukuran klasik, dengan menggunakan teori ini, sebuah tes dinilai berdasarkan masing-masing item. Sehingga setiap item memiliki tingkat kesukaran yang berbeda-beda, memperhitungkan bagaimana tingkat kemampuan peserta didik, serta karakteristik tes tidak bergantung terhadap responden.

Pada beberapa penelitian, penggunaan kedua jenis analisis tes tersebut dapat

digabungkan seperti pada penelitian (Ramalis, T. R., & Rusdiana, D. 2015) yang menganalisis karakteristik tes keterampilan berpikir kritis materi bumi dan antariksa yang diujikan pada calon guru. Namun pada studi penelitian ini, hanya akan focus membahas mengenai teori respon butir.

Penggunaan Teori Respon Butir sudah mulai banyak digunakan dalam menganalisis instrumen tes, termasuk dalam menganalisis instrumen tes materi fisika di sekolah salah satunya dalam mengukur dimensi kognitif (Mu'iz, M. S., Kaniawati, I., & Ramalis, T. R., 2018). Pada penelitian yang dilakukan oleh (Hakim, M. L., Muslim, M., & Ramalis, T. R., 2019) terkait penentuan karakteristik tes ranah kognitif pada materi elastisitas menggunakan teori respon butir dan menghasilkan tes yang memiliki validasi tinggi serta tes berkategori baik. Lalu penelitian yang dilakukan oleh (Rosidah, N. A., Ramalis, T. R., & Suyana, I., 2018) terkait karakteristik tes keterampilan berpikir kritis materi suhu dan kalor menggunakan teori respon butir yang menunjukkan bahwa tes tersebut dapat digunakan untuk mengukur kemampuan siswa pada tingkat kemampuan rendah sampai tinggi. Dan juga (Alifa, T. F., Ramalis, T. R., & Purwana, U. 2018) yang meneliti tentang penggunaan teori respon butir dalam menganalisis karakteristik tes penalaran ilmiah materi mekanika dan menunjukkan bahwa tes yang dibuat memiliki karakteristik tes yang baik serta reliabel untuk mengukur siswa dengan tingkat kemampuan sedang.

2. Kajian Literatur

a. Apa itu Teori Respon Butir?

Terdapat dua teori analisis yang saat ini berkembang dan banyak digunakan dalam merancang dan menganalisis suatu instrumen tes. Teori yang pertama yaitu teori tes klasik yang sudah banyak digunakan secara luas. Teori tersebut mulai berkembang sejak tahun 1940, namun teori ini memiliki keterbatasan dengan adanya sifat *group dependent* dan *item dependent* sehingga muncul teori kedua yang mampu mengatasi keterbatasan tersebut yaitu teori tes modern atau sering disebut juga dengan teori respon butir. Teori respon butir sudah mulai muncul berbarengan dengan teori tes klasik yaitu tahun 1946, namun ketertarikan

pada teori ini berkurang pada tahun 1960 karena adanya perkembangan teori *true score*. Namun pada saat teori *true score* mulai menarik perhatian, masalah dan kelemahan teori *true score* mulai mendapatkan perhatian yang akhirnya menyebabkan ketertarikan terhadap teori respon butir kembali meningkat. Teori respon butir menggunakan lebih banyak asumsi dari pada teori tes klasik, sehingga diharapkan dapat memberikan informasi yang lebih teliti.

Parameter yang digunakan dalam teori tes klasik dan teori respon butir pada dasarnya adalah sama yaitu tingkat kesukaran, daya pembeda dan factor tebakan semu (*pseudo guessing*). Perbedaannya terletak pada formula, skala, dan satuan yang digunakan. Kesalahan pengukuran menurut teori tes klasik dinyatakan dengan kesalahan baku pengukuran atau SEM (*Standar Error of Measurement*) yang besarnya bergantung pada indeks kehandalan tes. Sedangkan pada teori respon butir, kesalahan pengukuran dinyatakan dengan kesalahan baku pengukuran atau SEM (*Standar Error of Measurement*) yang besarnya bergantung pada tingkat kemampuan seseorang dan fungsi informasi tes.

b. Keunggulan Teori Respon Butir

Dengan menggunakan teori Respon Butir, ukuran taraf kesukaran butir serta ciri butir lainnya akan tetap (invarian) terhadap kelompok peserta tes, tidak masalah kelompok peserta mana saja yang mengerjakannya selama mereka mempunyai kemampuan yang memadai untuk mengerjakannya; dan dengan teori ini pula, ukuran kemampuan peserta akan tetap (invarian) terhadap kelompok butir tes, tidak masalah kelompok butir mana saja yang mereka kerjakan selama kelompok butir itu mampu secara memadai dikerjakan oleh peserta tes (Dali S.Naga, 1992). Berikut adalah kelebihan *Item Response Theory* (IRT) dalam (Hambleton, R. K., & Jones, R.W., 1993) Statistik butir soal yang tidak bergantung pada kelompok peserta tes.

1. Skor mendeskripsikan kemampuan peserta tes dan tidak bergantung pada tingkat kesulitan tes.

2. Model tes yang dapat digunakan untuk menghubungkan butir soal dengan tingkat kemampuan peserta tes.
3. Model tes yang tidak memerlukan tes secara paralel dalam menentukan reliabilitas.

c. Asumsi Teori Respon Butir

Terdapat tiga asumsi yang mendasari teori respon butir (Hambleton, Swaminathan, & Roger, 1991) yaitu:

1. Unidimesi

Unidimensi memiliki arti bahwa setiap butir tes hanya mengukur satu kemampuan. Namun pada faktanya, asumsi unidimensi tidak dapat dilakukan secara tepat karena ada beberapa faktor kognitif lain yang mempengaruhi, seperti kemampuan kognitif, kepribadian dan faktor-faktor pelaksanaan tes, seperti kecemasan, motivasi, dan tebakan.

2. Invariansi Parameter

Asumsi ini berarti karakteristik butir tes tidak tergantung pada distribusi parameter orang (kemampuan responden) dan parameter butir tes (tingkat kesukaran, daya pembeda, dsb.). Parameter orang (kemampuan responden) tidak akan berubah hanya karena mengerjakan tes yang berbeda tingkat kesukarannya dan parameter butir tes (tingkat kesukaran, daya pembeda, dsb.) tidak akan berubah hanya karena diujikan pada kelompok responden yang berbeda tingkat kemampuannya. Invariansi parameter akan terbukti jika hasil estimasi kemampuan responden tidak berbeda walaupun tes yang dikerjakan memiliki tingkat kesukaran berbeda dan jika hasil estimasi parameter butir tes tidak berbeda walaupun diujikan pada responden yang berbeda tingkat kemampuannya.

3. Independensi Lokal

Asumsi independensi lokal dibagi menjadi dua yaitu independensi lokal terhadap respon responden dan independensi lokal terhadap butir tes. Independensi lokal terhadap respons responden memiliki arti bahwa jawaban responden tidak dipengaruhi oleh jawaban responden lain. Sedangkan independensi lokal terhadap butir tes memiliki arti bahwa jawaban responden pada sebuah butir tes tidak dipengaruhi oleh jawaban responden itu pada butir tes yang lain.

d. Model dalam Teori Respon Butir

Berdasarkan jenis data yang diperoleh, ada beberapa model teori respon butir yang dapat digunakan, yaitu:

1) Model Dikotomi

Terdapat tiga jenis model yang dapat digunakan untuk menganalisis tes dikotomi, yaitu model 1 parameter logistik (1PL), model 2 parameter logistik (2PL) dan model 3 parameter logistik (3PL) (Retnawati, 2014).

Model 1 parameter logistik adalah salah satu model IRT yang paling sering digunakan, model ini biasa disebut dengan model Rasch. Model Rasch hanya menggunakan satu parameter soal, yaitu parameter tingkat kesukaran soal (b), sedangkan untuk parameter lainnya seperti parameter daya pembeda soal (a) dianggap sama untuk semua butir dan tebakan semu (c) sama dengan nol. Model 2 PL ini melibatkan dua parameter diantaranya parameter tingkat kesukaran (b) serta parameter daya pembeda butir (a). Model 3PL merupakan model yang melibatkan tiga parameter yaitu parameter daya pembeda butir (a), parameter tingkat kesukaran (b), dan faktor tebakan (*guessing*) (c). Dengan adanya faktor tebakan semu pada model logistik tiga parameter, memungkinkan siswa yang memiliki kemampuan rendah mempunyai peluang menjawab butir soal dengan benar.

2) Model Politomi

Untuk data yang bersifat politomi, analisis butir tes IRT dapat menggunakan model *Partial Credit Model* (PCM), *Generalized Partial Credit Model* (GPCM), *Graded Response Model* (GRM), *Modified Graded Response Model* (M-GRM), atau *Rating Scale Model* (RSM) (Widhiarso, 2010).

Model *Partial Credit Model* (PCM) atau model kredit parsial merupakan pengembangan dari Model 1-PL atau model *Rasch* yang diterapkan pada data politomi. Ini berarti, PCM masih tergolong dalam keluarga model *Rasch*. Model *Rasch* butir dikotomi yang hanya berisi satu parameter butir tes kemudian dikembangkan dengan menjabarkan parameter butir tes menjadi beberapa kategori. Model PCM cocok untuk diterapkan pada tes prestasi, misalnya soal hitungan fisika yang membutuhkan tahap identifikasi permasalahan hingga solusi akhir (Widhiarso, 2010). Untuk model GPCM, hampir mirip dengan model PCM, hanya

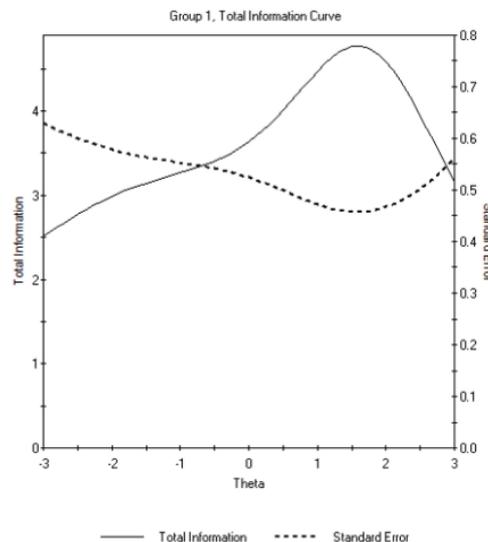
saja ada parameter kemiringan (*slope*) atau daya pembeda.

Model *Graded Response Model* (GRM) memiliki kemiripan dengan model 2-PL namun lebih spesifik, dalam setiap butir terdapat parameter b dalam setiap kategori kecuali yang pertama (DeMars, 2010). Dalam model ini, teknik penskoran memiliki banyak kategori yang tersusun dalam jawaban, seperti contohnya skala *likert* (Widhiarso, 2010). Untuk model M-GRM, masih sama seperti model GRM namun lebih terbatas karena mengkhendaki semua kategori memiliki jarak yang sama. Dalam M-GRM lokasi butir dan nilai parameter b dipisah sehingga jumlah parameternya lebih banyak dibanding GRM (Widhiarso, 2010).

e. Fungsi Informasi dan SEM (*Standard Error Measurement*)

Fungsi informasi merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat tes, pemilihan butir tes,

dan perbandingan beberapa perangkat tes (Retnawati, 2014). dalam IRT, kesalahan pengukuran atau biasa disebut *standard error measurement* (SEM) memiliki hubungan yang berbanding secara terbalik kuadratik dengan fungsi informasi, semakin tinggi nilai fungsi informasi maka SEM akan semakin kecil begitu pula sebaliknya (Retnawati, 2014). fungsi informasi digunakan untuk mengetahui reliabilitas dari instrumen tes yang dikonstruksi, semakin banyak informasi yang diperoleh maka semakin kecil SEM-nya. Perpotongan antara kurva fungsi informasi dengan SEM menunjukkan batas-batas estimasi reliabilitas dari tes tersebut. Fungsi informasi dapat direpresentasikan melalui sebuah kurva fungsi informasi yang menunjukkan hubungan antara variabel informasi pada sumbu y dan variabel kemampuan responden pada sumbu x . Berikut adalah salah satu contoh kurva fungsi informasi.



Gambar 1. Contoh Fungsi Informasi

f. *Item Characteristics Curves* (ICC) dan *Test Characteristics Curves* (TCC)

Dalam menentukan karakteristik tes seperti tingkat kesukaran (b) dan daya pembeda (a), digambarkan suatu hubungan antara peluang responden menjawab benar $P_i(\theta)$ dengan kemampuan responden (θ) melalui *Item Characteristics Curves* (ICC) dan *Test Characteristics Curves* (TCC).

Parameter b diartikan sebagai tingkat kesukaran yang merupakan suatu titik pada

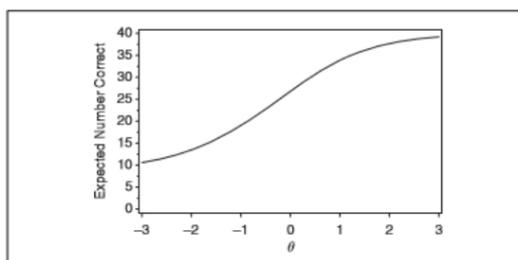
skala kemampuan agar peluang menjawab benar sebesar 50% artinya diperlukan kemampuan minimal yang sesuai dengan nilai parameter b yang diperoleh, untuk dapat menjawab benar dengan peluang 50% (Retnawati, 2014). Semakin besar nilai parameter b maka semakin besar kemampuan yang diperlukan untuk menjawab benar dengan peluang 50%.

Untuk parameter *slope* (a) diartikan sebagai daya pembeda yang dimiliki butir

ke-i. Dalam ICC, parameter a berupa kemiringan (*slope*) dari kurva titik b pada skala kemampuan tertentu (Retnawati, 2014), sehingga semakin besar kemiringannya, maka semakin besar pula daya pembeda dari butir tes tersebut.

Selain diinterpretasikan pada ICC, karakteristik tes secara keseluruhan dapat

dilihat melalui *Test Characteristics Curves* (TCC) (DeMars, 2010). Nilai terendah dalam sumbu Y menunjukkan skor terkecil yang dapat diraih oleh responden, sementara nilai tertinggi dalam sumbu Y menunjukkan skor terbesar yang dapat diraih oleh responden dari maksimal skor yang harusnya diperoleh.



Gambar 2. Contoh TCC

Parameter b pada TCC, dapat diperoleh dari hasil perpotongan antara perpanjangan nilai tengah TCC di sumbu Y dengan kurva TCC. Nilai tengah pada sumbu Y, dapat diperoleh dengan menjumlahkan skor terkecil dan skor terbesar yang ada pada sumbu Y, kemudian dibagi 2. Parameter b akan ditunjukkan oleh angka yang ada pada sumbu X. Sementara itu, untuk parameter a dapat diperoleh dengan menghitung $\tan a$ dari garis singgung yang dibuat pada TCC (DeMars, 2010).

g. Teori Respon Butir dalam menganalisis instrumen tes materi fisika

Berikut adalah kajian jurnal yang membahas mengenai penggunaan Teori Respon Butir dalam menganalisis instrumen tes materi fisika yang menunjukkan bahwa penggunaan Teori Respon Butir dapat memberikan gambaran rinci terkait instrumen tes yang digunakan:

- 1) Hasil analisis instrumen tes penalaran ilmiah materi suhu dan kalor pada jurnal (Willianti Saptawulan, dkk) yaitu tes memiliki karakteristik tes yang baik dan akan reliabel jika diberikan pada siswa dengan tingkat kemampuan sedang.
- 2) Penelitian yang dilakukan oleh (Rosidah, N. A., Ramalis, T. R., & Suyana, I., 2018) terkait karakteristik tes keterampilan berpikir kritis materi suhu dan kalor menggunakan teori respon butir yang menunjukkan bahwa tes

tersebut dapat digunakan untuk mengukur kemampuan siswa pada tingkat kemampuan rendah sampai tinggi.

- 3) (Alifa, T. F., Ramalis, T. R., & Purwana, U. 2018) yang meneliti tentang penggunaan teori respon butir dalam menganalisis karakteristik tes penalaran ilmiah materi mekanika dan menunjukkan bahwa tes yang dibuat memiliki karakteristik tes yang baik serta reliabel untuk mengukur siswa dengan tingkat kemampuan sedang.

3. Simpulan

Penelitian mengenai Teori Respon Butir atau *Items Response Theory* akan terus berkembang dalam menganalisis instrumen-instrumen tes terutama untuk materi fisika. Teori respon butir dapat lebih diunggulkan dari pada teori analisis modern karena beberapa asumsinya yang dapat menutupi kelemahan dari teori analisis klasik. Dengan menggunakan teori respon butir, karakteristik tes dapat diketahui secara lebih rinci untuk setiap butir soalnya.

Daftar Pustaka

- Alifa, T. F., Ramalis, T. R., & Purwana, U. 2018. Karakteristik Tes Penalaran Ilmiah Siswa SMA Materi Mekanika Berdasarkan Analisis Tes Teori Respon Butir. *Jurnal Inovasi dan Pembelajaran Fisika*, 5(1), 80-89.

- Amelia, R.N. & Kriswantoro. 2017. Implementasi *Item Response Theory* Sebagai Basis Analisis Kualitas Butir Soal dan Kemampuan Kimia Peserta didik Kota Yogyakarta. *Jurnal Kimia dan Pendidikan Kimia*, Vol. 2, No. 1, 1-12. DOI: 10.20961/jkpk.v2i1.8512.
- Arikunto, S. 2012. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT Bumi Aksara.
- Demars, C. 2010. *Item response theory*. New York : Oxford University Press.
- Hakim, M. L., Muslim, M., & Ramalis, T. R. 2019. Karakteristik Tes Hasil Belajar Ranah Kognitif Materi Elastisitas Menggunakan Analisis Item Response Theory. *Jurnal Penelitian Pembelajaran Fisika*, 10(1), 22-32.
- Hambleton, R.K. & Jones, R.W. 1993. Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. 1991. *Fundamental of Item Response Theory*. Newbury Park, CA: Sage Publication Inc.
- Mu'iz, M. S., Kaniawati, I., & Ramalis, T. R. 2018. Analyzing instrument characteristics of critical thinking skills and mastery of concepts based on item response theory. In *International Conference on Mathematics and Science Education of Universitas Pendidikan Indonesia* (Vol. 3, pp. 162-167).
- Naga, D.S. 1992. *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta : Gunadharma.
- Ramalis, T. R., & Rusdiana, D. 2015. Karakteristik Pengembangan Tes Keterampilan Berpikir Kritis Bumi dan Antariksa Untuk Calon Guru. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 1(2), 51-58.
- Retnawati, H. 2014. *Teori Respons Butir dan Penerapannya*. Yogyakarta: Nuha Medika.
- Rosidah, N. A., Ramalis, T. R., & Suyana, I. 2018. Karakteristik Tes Keterampilan Berpikir Kritis Berdasarkan Teori Respon Butir. *Jurnal Inovasi dan Pembelajaran Fisika*, 5(1), 54-63.
- Saptawulan, W., T.R Ramalis & Muslim 2018. *Karakterisasi Tes Penalaran Ilmiah Materi Suhu Dan Kalor Berdasarkan Teori Respon Butir*(Doctoral dissertation, Universitas Pendidikan Indonesia).
- Sutiadi, A., & Kurniawati, R. 2015. Analisis Butir Soal Ujian Nasional SMA Bidang Fisika Tahun 2014 Menggunakan Taxonomy of Introductory Physics Problem. *Prosiding Pertemuan Ilmiah XXIX HFI Jateng & DIY*.
- Widhiarso, W. 2010. *Model Politomi dalam Teori Respons Butir*. Yogyakarta: Psikologi UGM