

MENDETEKSI KEAJEGAN BUTIR TES DENGAN FUNGSI INFORMASI

Rahmadhani Mulvia, Taufik Ramlan Ramalis dan Ridwan Efendi

Universitas Pendidikan Indonesia , Bandung

Email: rahmadhanimulvia@upi.edu, taufik_lab.ipba@upi.edu dan
ridwanefendi@upi.edu

INFO ARTIKEL	ABSTRACT
Diterima 27 Desember 2020 Diterima dalam bentuk review 10 Januari 2021 Diterima dalam bentuk revisi 20 Januari 2021	<i>This study aims to detect the consistency of the test items with the information function of the item response theory. The accuracy of the information in measurement is known as reliability. The research method used is development with ADDIE design (Analyze, Design, Development, Implement, and Evaluation). Participants in this study consisted of 5 experts and 280 high school students who were selected using random sampling techniques. The instruments used were a validation sheet and a questionnaire for scientific thinking habits. The validation sheet is used at the development stage which consists of 6 aspects including construct, content and language. The questionnaire for scientific thinking habits is used at the trial stage which consists of 22 test items and is declared worthy of being tested by experts. The data obtained from the validation sheet were analyzed using the Many Facet Rasch Model assisted by the Minifac program from Winstep so that it was declared feasible to be tested. Data from the trial phase were analyzed using a graded response model assisted by the Eirt program. The results obtained are most of the good instruments used to measure scientific thinking habits and can measure the level of ability from low to high levels. Therefore, most of the test items can be implicated in measuring the ability of students to have scientific thinking habits.</i>
Keywords: information functions, reliability, item response theory	
Kata kunci: fungsi informasi, reliabilitas, teori respon butir.	ABSTRAK Penelitian ini bertujuan untuk mendeteksi keajegan butir tes dengan fungsi informasi dari teori respon butir. Keajegan informasi dalam pengukuran dikenal dengan reliabilitas. Metode penelitian yang digunakan adalah pengembangan dengan desain ADDIE (Analyze, Design, Development, Implement, and Evaluation). Partisipan pada penelitian ini terdiri dari 5 ahli dan 280 peserta didik SMA yang dipilih dengan menggunakan teknik random sampling. Instrumen yang digunakan adalah lembar validasi dan angket kebiasaan berpikir ilmiah. Lembar validasi digunakan pada tahap <i>development</i> yang terdiri dari 6 aspek mencakup konstruk, konten dan bahasa. Angket kebiasaan berpikir ilmiah digunakan pada tahap <i>implement</i> yang terdiri dari 22 butir tes dan dinyatakan layak untuk diujicobakan oleh ahli. Data yang diperoleh dari lembar validasi dianalisis dengan

menggunakan *Many Facet Rasch Model* berbantuan program Minifac dari Winstep sehingga dinyatakan layak untuk diuji cobakan. Data dari tahap *implement* dianalisis dengan menggunakan *graded response model* berbantuan program eirt. Hasil yang diperoleh adalah sebagian besar dari instrumen baik digunakan untuk mengukur kebiasaan berpikir ilmiah dan dapat mengukur tingkat kemampuan dari tingkat yang rendah sampai tingkat yang tinggi. Oleh karena itu, sebagian besar butir tes dapat diimplikasikan dalam pengukuran kemampuan kebiasaan berpikir ilmiah yang dimiliki oleh peserta didik.

Pendahuluan

Butir tes yang baik adalah butir yang memiliki tingkat keajegan yang stabil dalam memberikan informasi dari pengukuran yang dilakukannya. Istilah lain dikenal dengan tingkat reliabilitas. Untuk menentukan tingkat reliabilitas terdapat beberapa cara. Pada teori tes klasik, untuk data dikotomi, nilai reliabilitas ditentukan dengan teknik Kuder-Richardson (KR) yaitu KR-20 dan KR-21 sedangkan untuk data politomus, nilai reliabilitas ditentukan dengan *Spearman Brown*, *Alpha Cronbach*, *Koefisien Feldt*, *Theta Armor*, dan *Omega McDonald* (Sumintono & Widhiarso, 2014), (Arifin, 2014). Pada teori respon butir, nilai reliabilitas untuk data dikotomi dan politomi ditentukan dengan fungsi informasi (DeMars, 2010), (Retnawati, 2014), (Sumintono & Widhiarso, 2014).

Fungsi informasi adalah fungsi yang memberikan informasi tentang estimasi kemampuan responden dari model pada teori respon butir. Fungsi ini dapat menjelaskan untuk tingkat butir tes dan tingkat tes. Pada tingkat butir tes, fungsi informasi menyatakan keajegan atau kekuatan butir tes dalam menjelaskan kemampuan responden atau dikenal dengan latent trait yang diukur dengan tes (Myszkowski, 2019). Secara matematis, fungsi informasi untuk tingkat butir sebagai berikut:

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad \dots (1)$$

Keterangan :

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

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

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

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

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

Untuk tingkat tes, fungsi informasi tes merupakan jumlah dari fungsi informasi butir tes yang menyusun tes tersebut. Secara matematis, fungsi informasi tes sebagai berikut:

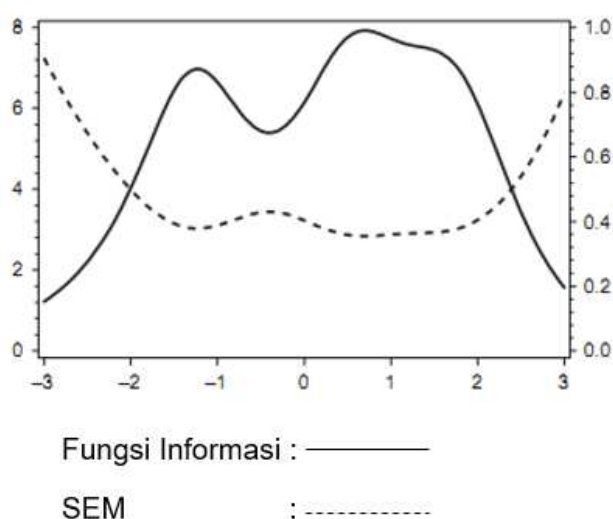
$$I(\theta) = \sum_{j=1}^n I_j \quad \dots (2)$$

Titik puncak dari fungsi informasi tidak selalu tunggal, (Sumintono & Widhiarso, 2014). Hal ini dapat disebabkan karena setiap butir tes memiliki perbedaan kemampuan

dalam mengukur. Semakin tinggi puncak informasi yang dapat dicapai, semakin banyak pula informasi yang dapat diperoleh untuk menjelaskan *latent trait* yang dimiliki oleh responden.

Informasi yang diperoleh dari hasil pengukuran tidak terlepas dari kesalahan. Kesalahan pengukuran dalam teori respon butir dikenal dengan *Standard Error of Measurement* (SEM). Hubungan SEM dan fungsi informasi adalah berbanding terbalik kuadrat sehingga semakin besar fungsi informasi maka SEM semakin kecil atau sebaliknya (Ningsih *et al.*, 2018), seperti Gambar 1). Secara matematis, SEM dapat dirumuskan sebagai berikut:

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad \dots (3)$$



Gambar 1
Fungsi Informasi dan SEM

Fungsi informasi dalam menunjukkan nilai reliabilitas bersifat independen terhadap yang lainnya (Tjun *et al.*, 2012). Hal ini dikarenakan fungsi informasi merupakan bagian dari hasil analisis dengan menggunakan teori respon butir yang dapat memperbaiki kekurangan dari teori tes klasik. Berikut kekurangan dari teori tes klasik yaitu: (1) reliabilitas bergantung pada karakteristik sampel yang diuji; (2) setiap butir tes bersifat *linear* tapi pada kenyataannya tidak selalu *linear*; (3) nilai *latent trait* yang diperoleh dan kemampuan; (4) konstruksi tes berubah dari waktu ke waktu; (5) hasil dari individu bergantung pada karakteristik butir yang diujikan (Huriaty, 2019), (Qasem, 2013), (Sumintono & Widhiarso, 2014), (Van Zile-Tamsen, 2017), (Sumintono & Widhiarso, 2015), (Eaton *et al.*, 2019). Kelemahan tersebut dapat diperbaiki oleh teori respon butir apabila terdapat kecocokan antara data hasil pengukuran dengan model yang digunakan untuk menganalisis pada teori respon butir (Retnawati, 2014).

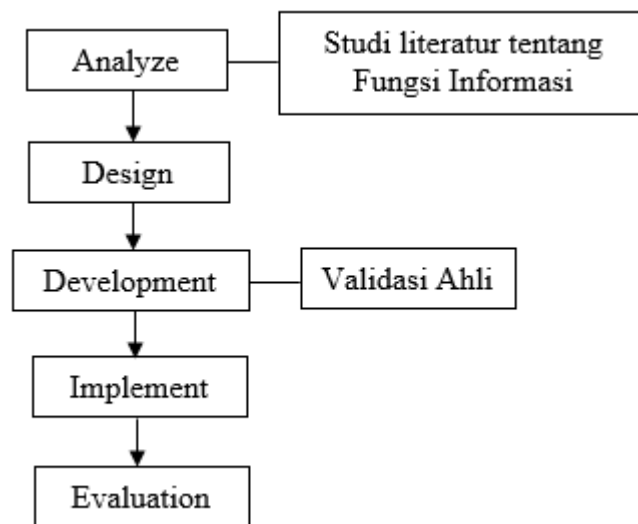
Dari pembahasan di atas, sesuatu yang dimiliki peserta didik akan *linear* tapi pada kenyataannya tidak selalu *linear*. Oleh karena itu, penelitian ini bertujuan untuk mendeteksi keajegan butir tes dalam pengukuran dengan meninjau fungsi informasi dari

analisis teori respon butir. Hal ini penting dilakukan karena pada penelitian yang telah ada untuk menentukan keajegan dari pengukuran atau reliabilitas hanya untuk tingkat tes dengan menggunakan teori tes klasik yang memiliki beberapa kelemahan.

Penelitian ini memiliki perbedaan dibandingkan penelitian lainnya yaitu dapat mendeteksi keajegan atau reliabilitas dari informasi hasil pengukuran pada tingkat butir sehingga diketahui butir tes yang reliabel dan memiliki ketelitian yang baik dalam mengukur kemampuan siswa. Selain itu, penelitian ini juga dapat memberikan informasi tentang kemampuan siswa yang dapat terukur oleh setiap butir yang ditunjukkan oleh nilai logit kemampuan siswa (θ). Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi perbaikan dalam pengembangan instrumen yang mengukur kemampuan siswa sehingga diperoleh reliabilitas dan ketelitian pengukuran yang baik dalam mengukur kemampuan siswa dari tingkat rendah sampai tinggi.

Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode pengembangan, dengan desain ADDIE (*Analyze, Design, Development, Implement, and Evaluation*), seperti Gambar 2. Hal ini dikarenakan tahapan tersebut sudah dapat mencapai tujuan dari penelitian.



Gambar 2
Metode Penelitian

a. Partisipan

Partisipan yang terlibat pada penelitian ini adalah 5 orang ahli dalam pendidikan fisika dan 280 orang siswa SMA di Jawa Barat dan dipilih dengan menggunakan teknik *random sampling*. Ahli dilibatkan pada tahapan *development* untuk memvalidasi instrumen yang telah didesain. Lima orang ahli ini terdiri dari 3 orang dosen dan 2 orang guru fisika.

Siswa SMA dilibatkan pada tahapan *implement*. Jumlah tersebut memadai untuk melakukan evaluasi sifat psikometrik dari butir tes. Hal ini dikarenakan 50

orang partisipan memadai untuk mengevaluasi sifat psikometrik sesuatu (Ananda *et al.*, 2017) Pendapat lainnya, ukuran partisipan dalam mengembangkan instrumen tergantung pada jumlah butirnya seperti 10 butir diperlukan 200 partisipan, 25 butir diperlukan 250 partisipan, dan 90 butir diperlukan 400 partisipan (Yamtinah & Budiyo, 2015).

b. Instrumen

Instrumen yang digunakan pada penelitian ini adalah lembar validasi yang terdiri dari 6 aspek penilaian mencakup konstruk, konten dan bahasa dari instrumen yang dikembangkan. Validasi dilakukan untuk memperoleh instrumen yang layak sebelum diujicobakan pada tahap *implement*.

Selain itu, digunakan pula instrumen angket kebiasaan berpikir ilmiah untuk tingkat SMA pada tahap *implement* yang telah dinyatakan layak oleh ahli. Instrumen tersebut berisi 22 butir tes yang memiliki 4 pilihan jawaban dengan rincian butir tes yaitu: 4 butir untuk mengukur ketidakpercayaan terhadap argumen yang dikemukakan ahli, 2 butir untuk mengukur bersikap terbuka terhadap ide-ide baru, 3 butir untuk mengukur meragukan tentang suatu hal, 4 butir untuk mengukur memiliki alasan yang sistematis dan logis dalam mengemukakan sesuatu, 2 butir untuk mengukur tidak memiliki keberpihakan pada sesuatu, 4 butir untuk mengukur menanggukkan kepercayaan terhadap sesuatu yang belum jelas, dan 3 butir untuk memiliki rasa ingin tahu.

c. Teknik Analisis

Data yang diperoleh adalah data kuantitatif yang berasal dari tahap *development* dan tahap *implement*. Pada tahap *development*, dihasilkan data hasil validasi instrumen yang melibatkan 5 orang ahli yang memiliki latar pendidikan fisika berupa data ordinal. Data tersebut dianalisis dengan menggunakan *Many Facet Rasch Model* (MFRM) berbantuan Minifac dari Winstep sehingga diperoleh nilai validitas berdasarkan ahli dengan meninjau nilai *raw variance explained by measures*. Berikut kategori nilai validitas berdasarkan nilai unidimensional, Tabel 1.

Tabel 1
Kategori Validitas

Nilai	Kategori
$20\% \leq X \leq 40\%$	Cukup
$40\% < X \leq 60\%$	Baik
$X > 60\%$	Sangat Baik

Sumber: Sumintono dan Widhiarso, 2014

Dari tahap *implement*, data diperoleh berupa politomi sehingga dianalisis dengan *graded response model* (GRM) dari teori respon butir berbantuan program eirt yang kompatibel dengan *microsoft excel*. Hasil yang diperoleh salah satunya adalah fungsi informasi.

Untuk fungsi informasi, semakin tinggi titik puncak dari fungsi informasi butir tes maka semakin ajeg atau reliabel butir tes tersebut dalam melakukan pengukuran (Hutahaean, 2015) (Retnawati, 2014). Sebaliknya, untuk nilai SEM, semakin rendah

titik puncak SEM maka semakin kecil kesalahan yang terjadi dalam pengukuran oleh butir tes tersebut.

Hasil dan Pembahasan

Hasil dari penelitian menggunakan metode pengembangan dengan desain ADDIE (*Analyze, Design, Development, Implement, and Evaluation*), sebagai berikut:

Tahap *analyze* menghasilkan informasi terkait keajegan butir dalam melakukan pengukuran atau dikenal dengan reliabilitas (DeMars, 2010), (Retnawati, 2014). Nilai tersebut dapat ditunjukkan dengan berbagai cara sesuai dengan data yang diperoleh dan teori analisis yang digunakan. Salah satu cara untuk menentukan keajegan atau reliabilitas tersebut dengan menggunakan fungsi informasi dari teori respon butir. Fungsi informasi ini dapat menunjukkan keajegan atau reliabilitas dari tingkat butir tes dan tingkat tes secara menyeluruh.

Selanjutnya, dilakukan tahap *design* instrumen untuk mengukur kebiasaan berpikir ilmiah yang memiliki tujuh karakteristik yaitu ketidakpercayaan terhadap argumen yang dikemukakan ahli, bersikap terbuka terhadap ide-ide baru, meragukan tentang suatu hal, memiliki alasan yang sistematis dan logis dalam mengemukakan sesuatu, tidak memiliki keberpihakan pada sesuatu, menanggukuhkan kepercayaan terhadap sesuatu yang belum jelas, dan memiliki rasa ingin tahu (Taylor, 2013). Instrumen ini terdiri dari 22 butir dengan 4 pilihan jawaban yang menunjukkan tingkat kepercayaan responden yaitu sangat benar, benar, salah dan sangat salah. Instrumen yang telah didesain, dikembangkan melalui tahap *development*.

Pada tahap *development* dilakukan dengan melakukan validasi instrumen. Instrumen divalidasi oleh 5 ahli berlatar belakang fisika dengan menggunakan lembar validasi sehingga diperoleh data ordinal. Data tersebut dianalisis dengan Many Facet Rasch Model (MFRM) berbantuan Minifac dari Winstep sehingga diperoleh nilai validitas. Berdasarkan analisis tersebut, diperoleh nilai validitas yang ditunjukkan oleh *raw variance explained by measures* sebesar 27,42%. Nilai tersebut berada pada kategori cukup baik (Amalia & Susilaningih, 2014), artinya instrumen yang telah didesain dinyatakan layak untuk diujicobakan berdasarkan penilaian ahli.

Implement dilakukan pada 280 orang siswa SMA di Jawa Barat dengan menggunakan bantuan Google Form. Data yang diperoleh berupa data ordinal yang dianalisis dengan *graded response model* (GRM) dari teori respon butir berbantuan program eirt.

Data yang dianalisis dengan menggunakan GRM dievaluasi pada tahap *evaluation* sehingga dapat menghasilkan fungsi informasi dan SEM. Fungsi informasi butir tes dan fungsi informasi tes secara menyeluruh dapat diperoleh secara langsung dari keluaran analisis dengan program eirt. Akan tetapi, untuk nilai SEM butir tes dan SEM tes secara keseluruhan dapat dibuat secara manual dengan menggunakan persamaan 3.

Untuk memperoleh estimasi fungsi informasi dan SEM dari analisis GRM yang baik, maka harus ditinjau kesesuaian data dengan model berdasarkan nilai chi-kuadrat. Berikut ini nilai chi-kuadrat setiap butir berdasarkan hasil analisis GRM, Tabel 2.

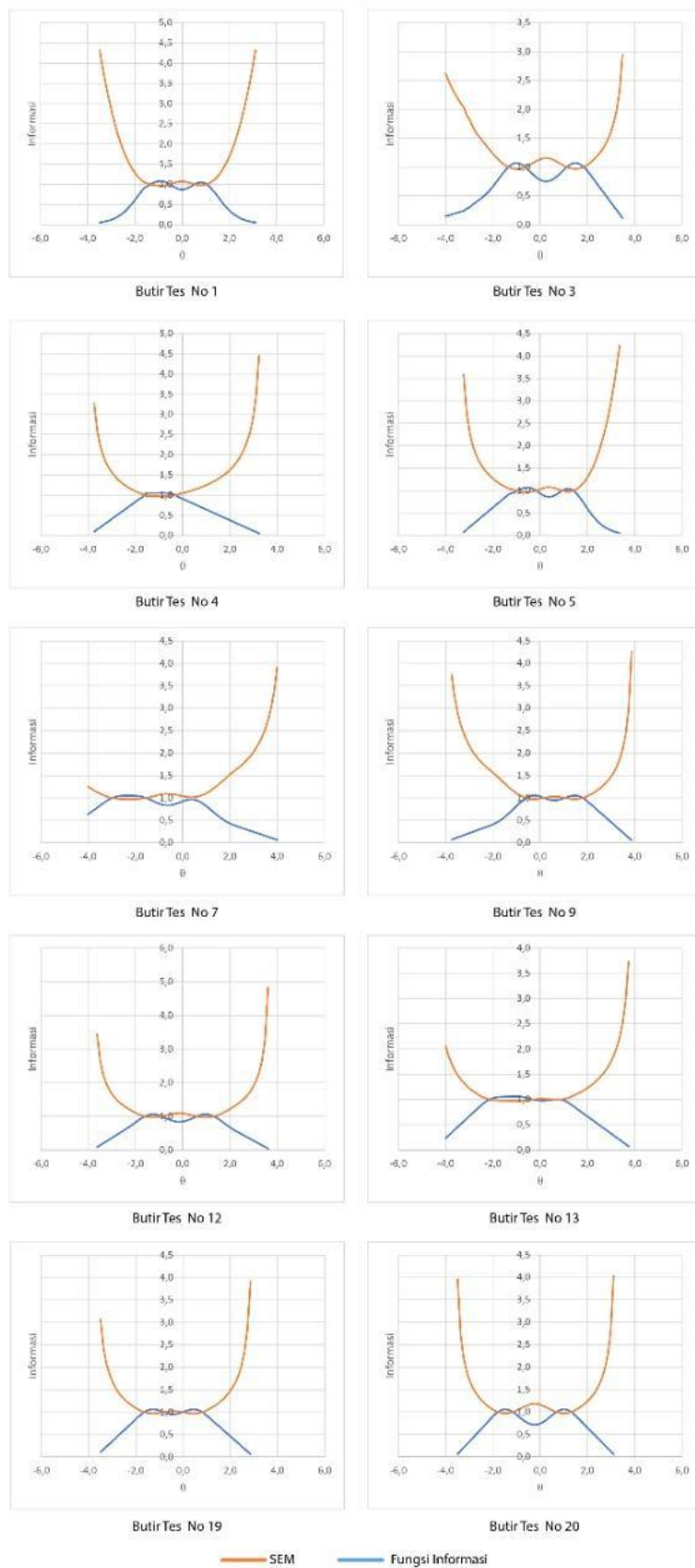
Tabel 2
Chi-Kuadrat Hasil Analisis dengan GRM

No	Chi-square	Derajat Kebebasan
1	16,345	20
2	6,907	20
3	8,313	20
4	8,824	30
5	7,553	30
6	9,914	20
7	16,494	30
8	9,833	30
9	1,485	20
10	18,089	30
11	6,996	20
12	8,417	20
13	9,360	30
14	6,996	20
15	9,210	30
16	8,149	20
17	10,754	20
18	8,316	20
19	1,485	20
20	8,308	20
21	9,313	20
22	16,345	20

Dari Tabel 2 diketahui bahwa setiap butir tes memiliki nilai chi-kuadrat hitung yang lebih kecil dibandingkan dengan chi-kuadrat tabel dengan derajat kebebasan tertentu. Oleh karena itu, dapat dinyatakan bahwa data setiap butir tes memiliki kesesuaian yang baik atau valid untuk dianalisis dengan menggunakan GRM sehingga menghasilkan estimasi fungsi informasi dan SEM yang baik (Fitriani et al., 2019).

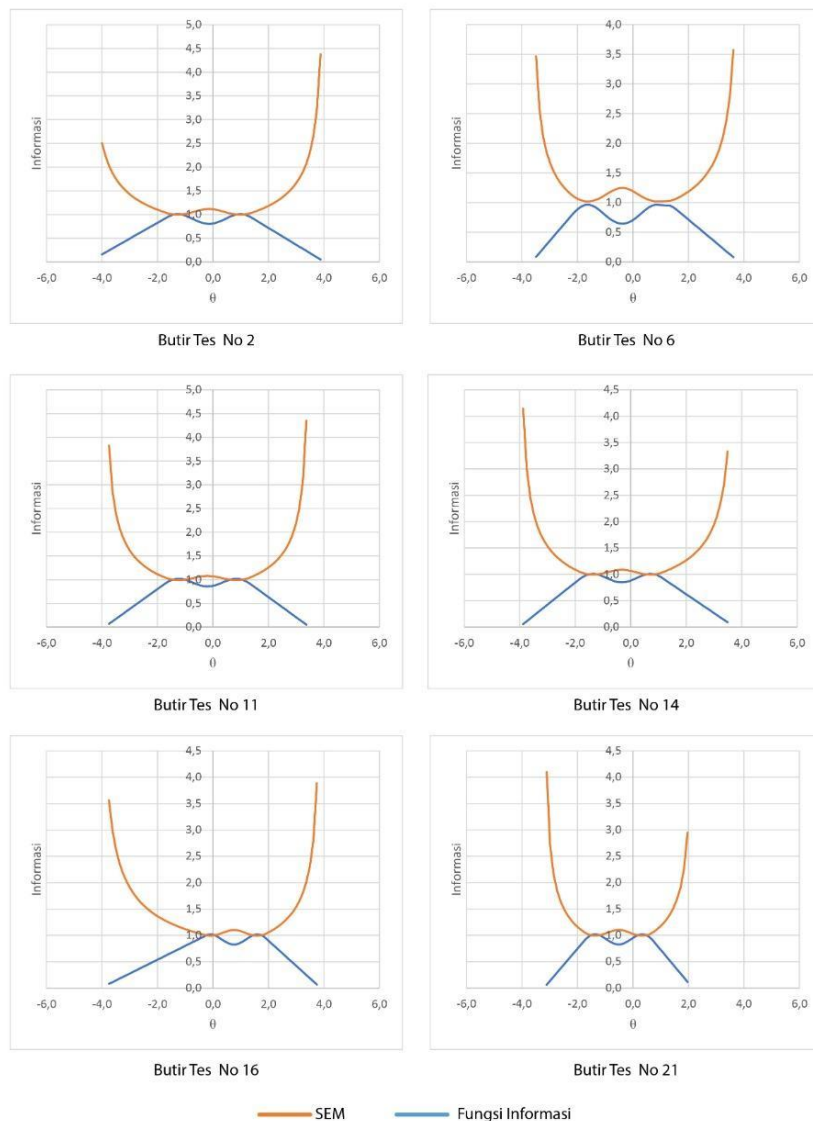
Selanjutnya dapat dilakukan analisis untuk menentukan fungsi informasi dan SEM untuk setiap butir tes. Dari tahap evaluasi diperoleh hasil seperti Gambar 3.

Mendeteksi Keajegan Butir Tes Dengan Fungsi Informasi



Gambar 3
Fungsi Informasi dan SEM dari Butir Tes yang Baik

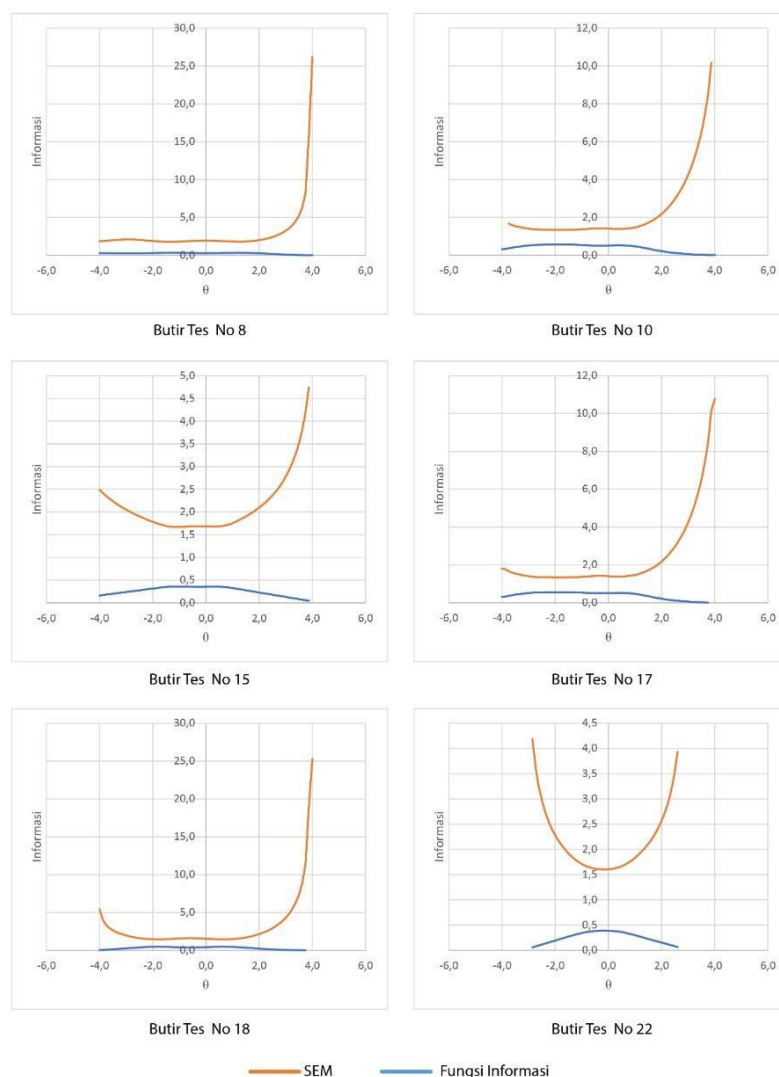
Gambar 3 menunjukkan bahwa 10 dari 22 butir tes memiliki nilai puncak informasi lebih tinggi dibandingkan dengan nilai puncak SEM. Artinya, butir tes tersebut memiliki nilai kejelasan atau reliabilitas dalam pengukuran yang baik sehingga informasi yang diberikan sebagai hasil pengukuran dapat dinyatakan stabil. Untuk nilai SEM yang diperoleh memiliki titik puncak yang lebih rendah dengan fungsi informasi sehingga dapat dinyatakan bahwa kesalahan dalam pengukuran rendah. Oleh karena itu, dapat dinyatakan bahwa butir tes tersebut baik digunakan dalam pengukuran. Selain itu, butir tes tersebut dapat mengukur peserta didik dari tingkat kemampuan yang rendah sampai kemampuan yang tinggi dengan nilai diantara $-4,0 < \theta < +4,0$.



Gambar 4
Fungsi Informasi dan SEM dari Butir Tes yang Cukup Baik

Dari Gambar 4, diketahui bahwa 6 dari 22 butir tes memiliki nilai titik puncak informasi yang sama dengan nilai titik puncak dari SEM. Artinya, butir tes tersebut

memiliki nilai keajegan atau reliabilitas dalam pengukuran yang baik tetapi memiliki kesalahan pengukuran yang cukup rendah. Oleh karena itu, dapat dinyatakan bahwa butir tes tersebut cukup baik digunakan dalam pengukuran. Selain itu, butir tes tersebut dapat mengukur peserta didik dari tingkat kemampuan yang rendah sampai kemampuan yang tinggi dengan nilai diantara $-4,0 < \theta < +4,0$.



Gambar 5
Fungsi Informasi dan SEM yang Tidak Baik

Sedangkan, 6 butir lainnya memiliki fungsi informasi dan SEM seperti Gambar 5. Dari Gambar 5 ditunjukkan bahwa setiap butir tes memiliki titik puncak fungsi informasi yang lebih rendah daripada titik puncak dari SEM. Artinya, butir tersebut sedikit memberikan informasi yang ajeg atau reliabel dalam melakukan pengukuran. Selain itu, butir tes tersebut memiliki estimasi kesalahan pengukuran yang sangat tinggi. Oleh karena itu, dapat dinyatakan bahwa butir tes tersebut tidak baik digunakan dalam pengukuran.

Kesimpulan

Fungsi informasi dari teori respon butir dapat digunakan untuk mengetahui keajegan atau reliabilitas butir tes. Dari instrumen tes yang dikembangkan diperoleh hasil bahwa sebagian besar butir tes yang dikembangkan dapat digunakan untuk mengukur kebiasaan berpikir ilmiah yang dimiliki oleh siswa. Butir tes tersebut dapat mengukur tingkat kemampuan dari tingkat yang rendah sampai tingkat yang tinggi. Oleh karena itu, sebagian besar butir dari instrumen tersebut dapat diimplikasikan dalam pengukuran kemampuan kebiasaan berpikir ilmiah yang dimiliki oleh siswa.

Bibliografi

- Amalia, N. F., & Susilaningsih, E. (2014). Pengembangan instrumen penilaian keterampilan berpikir kritis siswa sma pada materi asam basa. *Jurnal Inovasi Pendidikan Kimia*, 8(2).
- Ananda, R., Rafida, T., & Wijaya, C. (2017). *pengantar evaluasi program Pendidikan*. Perdana.
- Arifin, Z. (2014). *Evaluasi pembelajaran*. Bandung: PT Remaja Rosdakarya.
- DeMars, C. (2010). *Item response theory*. Oxford University Press.
- Eaton, P., Johnson, K., Frank, B., & Willoughby, S. (2019). Classical test theory and item response theory comparison of the brief electricity and magnetism assessment and the conceptual survey of electricity and magnetism. *Physical Review Physics Education Research*, 15(1), 10102.
- Fitriani, L., Ramalis, T. R., & Efendi, R. (2019). Karakterisasi Tes Keterampilan Proses Sains Materi Fluida Statis Berdasarkan Teori Respon Butir. *Omega: Jurnal Fisika Dan Pendidikan Fisika*, 5(2), 27.
- Huriaty, D. (2019). Analisis Karakteristik Parameter Butir Berdasarkan Model Logistik 3 Parameter. *Lentera: Jurnal Pendidikan*, 14(2), 33–40.
- Hutahaeen, J. (2015). *Konsep sistem informasi*. Deepublish.
- Myszkowski, N. (2019). Development of the R library “jrt”: Automated item response theory procedures for judgment data and their application with the consensual assessment technique. *Psychology of Aesthetics, Creativity, and the Arts*.
- Ningsih, D. R., Ramalis, T. R., & Purwana, U. (2018). Pengembangan Tes Keterampilan Berpikir Kritis Berdasarkan Analisis Teori Respon Butir. *WaPFI (Wahana Pendidikan Fisika)*, 3(2), 45–50.
- Qasem, M. A. N. (2013). A comparative study of classical theory (CT) and item response theory (IRT) in relation to various approaches of evaluating the validity and reliability of research tools. *IOSR Journal of Research & Method in Education (IOSR-JRME)*, 3(5), 77–81.
- Retnawati, H. (2014). Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana. *Yogyakarta: Nuha Medika*.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Trim Komunikata Publishing House.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Trim komunikata.

- Taylor, R. (2013). *Kiat-kiat Pede Untuk Meningkatkan Rasa Percaya Diri*. Gramedia Pustaka Utama.
- Tjun, L. T., Marpaung, E. I., & Setiawan, E. (2012). *Pengaruh kompetensi dan independensi auditor terhadap kualitas audit*.
- Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. *Research in Higher Education*, 58(8), 922–933.
- Yamtinah, S., & Budiyo, B. (2015). Pengembangan instrumen diagnosis kesulitan belajar pada pembelajaran kimia di SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(1), 69–81.