



KARAKTERISTIK PERANGKAT TES PRESTASI DAN KEMAMPUAN SISWA SEKOLAH DASAR BERDASARKAN TEORI RESPON BUTIR

Zulfa Amirul Mu'adzah*, Universitas Negeri Yogyakarta

Heri Retnawati, Universitas Negeri Yogyakarta

*e-mail: zulfaamirul.2019@student.uny.ac.id

Abstrak. Penelitian ini merupakan studi kasus dengan pendekatan kuantitatif untuk mengetahui karakteristik perangkat tes prestasi dan kemampuan siswa sekolah dasar berdasarkan teori respon butir. Penelitian ini bertujuan untuk: (1) Mengetahui model yang paling cocok dengan perangkat tes dan prestasi siswa sekolah dasar berdasarkan teori respon butir; (2) Mengetahui karakteristik butir-butir soal yang dengan model yang paling cocok berdasarkan teori respon butir. Data penelitian berupa hasil pengerjaan sebanyak 329 siswa terhadap soal penilaian harian mata pelajaran Matematika di sekolah dasar. Peneliti menggunakan data sekunder berupa data dikotomi yang diperoleh dari suatu perguruan tinggi di Yogyakarta. Estimasi parameter butir dilakukan dengan menggunakan model paling cocok dengan data berdasarkan teori respon butir. Model yang digunakan adalah model Rasch, 1PL, 2PL, 3PL dan 4PL. Penelitian ini harus memenuhi beberapa asumsi yakni asumsi unidimensi, invariansi parameter, dan independensi lokal. Hasil penelitian menunjukkan bahwa model yang paling cocok dalam penelitian ini adalah model 2PL. Model tersebut merupakan model yang terbaik diantara model Rasch, 1PL, 3PL, dan 4PL untuk mengestimasi data perangkat tes prestasi dan kemampuan siswa sekolah dasar berdasarkan teori respon butir. Berdasarkan nilai daya beda (a) dan tingkat kesulitan (b) didapatkan 12 butir soal dengan kualitas baik dan 8 butir soal dengan kualitas kurang baik.

Kata kunci: butir soal, karakteristik perangkat tes, teori respon butir.

Abstract. This research is a case study with a quantitative approach to determine the characteristics of achievement test equipment and abilities of elementary school students based on item response theory. This research aims to: (1) Find out the model that best suits the test equipment and elementary school student achievement based on item response theory; (2) Knowing the characteristics of the items with the most suitable model based on item response theory. The research data consists of the results of 329 students' work on daily assessment questions for Mathematics subjects in elementary schools. Researchers used secondary data in the form of dichotomous data obtained from a university in Yogyakarta. Item parameter estimation was carried out using the model that best fits the data based on item response theory. The models used are the Rasch model, 1PL, 2PL, 3PL and 4PL. This research must meet several assumptions, namely unidimensional assumptions, parameter invariance, and local independence. The research results show that the most suitable model in this research is the 2PL model. This model is the best model among the Rasch, 1PL, 3PL, and 4PL models for estimating achievement and ability test device data for elementary school students based on

item response theory. Based on the difference power value (a) and level of difficulty (b), 12 questions were obtained with good quality and 8 questions with poor quality.

Keywords: *test items, test equipment characteristics, item response theory.*

PENDAHULUAN

Peran pendidikan dalam pembangunan dan kemajuan Negara Republik Indonesia sangatlah penting. Perlunya evaluasi-evaluasi dalam dunia pendidikan harus ada dalam setiap instansi pendidikan untuk menaikkan standar mutu pendidikan Indonesia. Evaluasi mutu pendidikan dapat dilakukan dengan membaginya ke beberapa bidang, salah satunya adalah melalui penilaian hasil belajar siswa. Menurut Sudjana (2010) metode penilaian terbagi menjadi metode tes dan metode non tes. Salah satu langkah dalam mengevaluasi dengan metode penilaian tes pada perangkat tes adalah dengan melihat karakteristik butir-butir perangkat tes yang digunakan. Karakteristik butir soal atau perangkat tes dapat dilihat melalui 2 pendekatan yang sering digunakan, yakni *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT). CTT atau yang biasa disebut dengan teori tes klasik merupakan teori yang cukup mudah dalam penggunaannya, namun CTT memiliki beberapa kelemahan yakni estimasi tingkat kesulitan (b) dan daya beda (a) tergantung pada peserta tes, perbandingannya dilakukan melalui peserta yang dikelompok-kelompokkan, *standard error of measurement* (SEM) berlaku pada seluruh peserta tes. Model alternatif untuk mengatasi kelemahan pada teori tes klasik adalah dengan menggunakan teori tes modern atau salah satunya biasa disebut dengan teori respon butir (*item response theory*). Teori respon butir merupakan cara yang dapat mengestimasi parameter-parameter dalam sebuah model (Alfarisa & Purnama, 2019).

Mayoritas peneliti sebelumnya banyak menggunakan teori tes klasik dan perangkat tes yang digunakan adalah perangkat tes pada penilaian akhir sekolah. Dalam penelitian ini akan dianalisis perangkat tes prestasi yang berupa perangkat tes pada penilaian harian mata pelajaran Matematika berdasarkan teori respon butir. Teori respon butir memiliki beberapa model yang sering digunakan. Menurut Retnawati (2014), analisis dengan teori respon butir dapat menggunakan tiga model yakni model 1PL (1 parameter logistik) atau Rasch, model 2PL, dan model 3PL. Hambleton (1993), mengungkapkan bahwa IRT terdiri dari empat model yakni model 1PL, model 2PL, model 3PL, dan model 4PL. Model Rasch dan 1 PL hanya mengestimasi tingkat kesulitan butir saja. Model 2PL mengestimasi daya beda dan tingkat kesulitan. Model 3PL mengestimasi daya beda, tingkat kesulitan, dan tebakan semu, sedangkan model 4PL mengestimasi daya beda, tingkat kesulitan, tebakan semu, dan kecerobohan (*carelessness*). Pada penelitian ini akan digunakan data dikotomi terkait jawaban siswa terhadap penilaian harian mata pelajaran Matematika pada materi pecahan di sekolah dasar. Data penelitian akan diestimasi dengan menggunakan model Rasch, 1PL, 2PL, 3PL, dan 4PL. Hasil perhitungan masing-masing model akan dibandingkan berdasarkan jumlah butir yang cocok (*fit*) dengan model tersebut. Model dengan jumlah butir cocok paling banyak akan digunakan dalam penelitian.

Mengingat evaluasi perangkat tes sangatlah penting, peneliti mencoba mengidentifikasi karakteristik butir-butir soal dengan model yang paling cocok berdasarkan teori respon butir. Adanya penelitian ini bertujuan untuk: (1) Mengetahui model yang paling cocok dengan perangkat tes dan prestasi siswa sekolah dasar berdasarkan teori respon butir; (2) Mengetahui karakteristik butir-butir soal yang dengan model yang paling cocok berdasarkan teori respon butir. Butir-butir soal yang paling cocok tersebut akan diestimasi karakteristik butir soalnya. Butir-butir soal yang masih kurang baik untuk diujikan perlu diubah atau dilakukan penggantian dengan soal yang lebih baik. Hal tersebut dilakukan agar perangkat tes dapat benar-

benar mengukur kemampuan siswa dan dapat menjadi alat pengukuran dengan standar yang baik.

METODE

Penelitian ini bertujuan untuk mengetahui kualitas dan karakteristik perangkat tes dan prestasi siswa sekolah dasar berdasarkan teori respon butir. Penelitian ini merupakan studi kasus dengan pendekatan kuantitatif. Menurut Sugiono (2010) penelitian kuantitatif merupakan penelitian yang menganalisis nilai-nilai dan bisa dihitung. Penelitian ini juga termasuk dalam penelitian deskriptif eksploratif dengan menggunakan hasil tes yang diujikan kepada peserta tes. Penelitian ini menggunakan data hasil pengerjaan siswa pada penilaian harian mata pelajaran Matematika pada materi pecahan di 12 sekolah dasar yang tersebar di 7 provinsi berbeda di Indonesia. Data penelitian ini menggunakan seluruh responden yang ada. Data sebanyak 329 responden diurutkan dan diberi nomor responden ke-1 hingga responden ke-329. Penelitian ini menggunakan data sekunder yang pengumpulan datanya dilakukan oleh sekelompok mahasiswa pada salah satu universitas di Yogyakarta dan didapatkan oleh peneliti melalui koordinasi dengan dosen pembimbing.

Langkah-langkah dalam penelitian ini meliputi (1) mempersiapkan data berupa hasil pengerjaan siswa terhadap perangkat tes penilaian harian materi pecahan; (2) memastikan data yang akan digunakan dalam penelitian; (3) menguji kecocokan model dengan uji kecocokan model dan diperoleh model yang terbaik, uji kecocokan model dalam penelitian ini mengestimasi model Rasch, 1PL, 2PL, 3PL, dan 4PL, model dipilih berdasarkan jumlah butir yang cocok paling banyak pada masing-masing model; (4) melakukan uji asumsi berdasarkan teori respon butir. Uji asumsi tersebut yaitu meliputi uji asumsi unidimensi, uji asumsi independensi lokal, dan uji asumsi invariansi parameter; (5) melakukan estimasi parameter butir yang berupa karakteristik perangkat tes dan menghasilkan parameter daya beda (a), tingkat kesukaran (b), dan mengestimasi parameter kemampuan berdasarkan model terbaik.

Teknik Analisis Data

Penelitian ini akan mengestimasi karakteristik butir soal berdasarkan teori respon butir. Sebelum melakukan estimasi tersebut, perlu dilakukan uji kecocokan model berdasarkan teori respon butir. Model dalam teori respon butir adalah sebagai berikut:

1. Model Rasch

Model Rasch mengestimasi satu karakteristik butir soal saja yakni tingkat kesulitan (*difficulty*).

$$P(X_{ij} = 1 | \theta_j) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \quad (1)$$

$i = 1, 2, 3, \dots, m$, $j = 1, 2, 3, \dots, n$ dengan m adalah banyaknya butir tes, n adalah banyaknya peserta tes, $P(X_{ij} = 1)$ adalah probabilitas jawaban peserta tes ke- j menjawab benar pada butir ke- i , b_i adalah parameter kesulitan butir ke- i , e adalah konstanta dengan nilai berkisar 2,718, dan θ_j adalah parameter kemampuan peserta ke- j .

2. Model 1PL

Model 1PL hanya mengestimasi satu karakteristik butir soal, yakni tingkat kesulitan (*difficulty*).

$$P(X_{ij} = 1 | \theta_j) = \frac{e^{(Da\{\theta_j - b_i\})}}{1 + e^{a\{\theta_j - b_i\}}} \quad (2)$$

$i = 1, 2, 3, \dots, m, j = 1, 2, 3, \dots, n$ dengan m adalah banyaknya butir tes, n adalah banyaknya peserta tes, $P(X_{ij} = 1)$ adalah probabilitas jawaban peserta tes ke- j menjawab benar pada butir ke- i , b_i adalah parameter kesulitan butir ke- i , e adalah konstanta dengan nilai berkisar 2,718, D adalah konstanta dengan nilai berkisar 1,7, θ_j adalah parameter kemampuan peserta ke- j , dan a adalah parameter daya beda butir untuk semua butir (*discrimination*).

3. Model 2PL

Model 2PL mengestimasi karakteristik butir soal tingkat kesulitan (*difficulty*) dan daya beda (*discrimination*).

$$P(X_{ij} = 1|\theta_j) = \frac{e^{(a_i\{\theta_j-b_i\})}}{1 + e^{a_i\{\theta_j-b_i\}}} \quad (3)$$

$i = 1, 2, 3, \dots, m, j = 1, 2, 3, \dots, n$, dengan m adalah banyaknya butir tes, n adalah banyaknya peserta tes, $P(X_{ij} = 1)$ adalah probabilitas jawaban peserta tes ke- j menjawab benar pada butir ke- i , b_i adalah parameter kesulitan butir ke- i , e adalah konstanta dengan nilai berkisar 2,718, D adalah konstanta dengan nilai berkisar 1,7, θ_j adalah parameter kemampuan peserta ke- j , dan a adalah parameter daya beda butir untuk semua butir (*discrimination*).

4. Model 3PL

Model 3PL mengestimasi karakteristik butir soal tingkat kesulitan (*difficulty*), daya beda (*discrimination*), dan tebakan semu (*pseudo guessing*).

$$P(X_{IJ} = 1|\theta_j) = c_i + (1 - c_i) \frac{e^{(a_i\{\theta_j-b_i\})}}{1 + e^{a_i\{\theta_j-b_i\}}} \quad (4)$$

$i = 1, 2, 3, \dots, m, j = 1, 2, 3, \dots, n$ dengan m adalah banyaknya butir tes, n adalah banyaknya peserta tes, $P(X_{ij} = 1)$ adalah probabilitas jawaban peserta tes ke- j menjawab benar pada butir ke- i , b_i adalah parameter kesulitan butir ke- i , e adalah konstanta dengan nilai berkisar 2,718, θ_j adalah parameter kemampuan peserta ke- j , a adalah parameter daya beda butir untuk semua butir (*discrimination*), c_i adalah parameter tebakan semu (*pseudo guessing*) butir ke- i .

5. Model 4PL

Model 4PL mengestimasi karakteristik butir soal tingkat kesulitan (*difficulty*), daya beda (*discrimination*), tebakan semu (*pseudo guessing*), dan *carelessness*. Model 4PL merupakan pengembangan dari model 3PL yang menambahkan parameter *upper asymptote* pada model 3PL.

$$P(X_{ij} = 1|\theta_j) = g_i + (u_i - g_i) \frac{e^{(a_i\{\theta_j-b_i\})}}{1 + e^{a_i\{\theta_j-b_i\}}} \quad (5)$$

$i = 1, 2, 3, \dots, m, j = 1, 2, 3, \dots, n$ dengan m adalah banyaknya butir tes, n adalah banyaknya peserta tes, $P(X_{ij} = 1)$ adalah probabilitas jawaban peserta tes ke- j menjawab benar pada butir ke- i , b_i adalah parameter kesulitan butir ke- i , e adalah konstanta dengan nilai berkisar 2,718, θ_j adalah parameter kemampuan peserta ke- j , a adalah parameter daya beda butir untuk semua butir (*discrimination*), u_i adalah batas atas atau nilai maksimum peluang siswa menjawab benar sepanjang skala θ , g_i adalah parameter tebakan semu (*pseudo guessing*) butir ke- i .

Model terbaik pada penelitian ini didapat berdasarkan banyaknya butir yang cocok dengan model. Model dengan butir cocok paling banyak merupakan model terbaik atau model yang akan dipakai dalam penelitian ini. Setelah mendapatkan model yang cocok, maka harus diuji asumsi berdasarkan teori respon butir. Pengujian ini dilakukan apakah data yang akan dipakai dapat memenuhi asumsi-asumsi dalam teori respon butir atau belum. Asumsi-asumsi teori respon butir adalah sebagai berikut:

1. Unidimensi

Asumsi unidimensi memiliki arti bahwa setiap butir tes hanya mengukur satu kemampuan saja. Apabila suatu soal tes mengukur lebih dari satu dimensi, jawaban butir tersebut merupakan kombinasi dari beberapa kemampuan peserta tes. Asumsi unidimensi dapat ditunjukkan dari plot nilai eigen (*eigenvalue*) yang menunjukkan satu komponen dominan (Hambleton et al., 1991). Syarat unidimensi dapat terpenuhi ketika *eigenvalue* salah satu faktor memiliki nilai yang lebih dominan dibandingkan dengan *eigenvalue* faktor yang lain. *Eigenvalue* dapat dicari melalui perhitungan matriks varian kovarian yang menghitung kovarians masing-masing kolom dan baris. *Eigenvalue* yang telah didapat dari matriks varian kovarian dapat dibuat *scree plot* dan dilakukan analisis *Principal Component Analysis* (PCA). Pada penelitian ini, peneliti menggunakan bantuan *software* RStudio sehingga analisis PCA dapat dilakukan menggunakan *function* *PCA()* dan *function* *get_eigenvalue()* pada *package* *factoextra*.

2. Invariansi Parameter

Asumsi invariansi parameter berarti karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak bergantung dari ciri butir soal (Retnawati, 2014). Setiap peserta memiliki kemampuan masing-masing dan kemampuan tersebut akan tetap dan tidak berubah ketika menghadapi soal dengan tingkat kesulitan yang berbeda. Menurut Lord (1990), invariansi parameter merupakan karakteristik yang paling penting dalam IRT. Hal tersebut dikarenakan perbedaan antara teori klasik dan teori modern terletak pada invariansi parameter, di mana penskoran pada teori modern *invariant* (tidak berubah) terhadap butir tes serta terhadap peserta tes. Asumsi invariansi parameter akan terbukti ketika hasil estimasi parameter butir tidak berbeda walaupun sudah diujikan kepada peserta tes dengan tingkat kemampuan yang berbeda.

3. Independensi Lokal

Asumsi independensi lokal dibagi menjadi dua, yaitu independensi lokal terhadap jawaban peserta tes dan independensi lokal terhadap butir soal tes (Allen & Yen, 1989). Independensi lokal terhadap jawaban peserta tes yakni benar atau salahnya jawaban yang diberikan oleh peserta tes tidak memengaruhi benar salahnya jawaban yang diberikan oleh peserta lain. Sedangkan independensi lokal terhadap butir soal tes yakni benar salahnya peserta tes memberikan jawaban terhadap satu butir soal tes tidak memengaruhi benar salahnya jawaban peserta tes terhadap butir soal yang lain. Dengan menggunakan RStudio dapat menggunakan *function* *cor()* untuk melihat nilai korelasinya. Asumsi independensi lokal memiliki persamaan matematis sebagai berikut:

$$\begin{aligned} P(U_1, U_2, \dots, U_n | \theta) &= P(U_1 | \theta) \times P(U_2 | \theta) \times \dots \times P(U_n | \theta) \\ &= \prod_{i=1}^n P(U_i | \theta) \end{aligned} \quad (6)$$

dengan n adalah banyaknya butir dalam perangkat tes, $P(U_i | \theta)$ adalah Probabilitas peserta tes yang memiliki kemampuan θ dapat menjawab benar pada butir ke- i , dan

$P(U_1, U_2, \dots, U_n | \theta)$ adalah probabilitas peserta tes yang memiliki kemampuan θ dapat menjawab benar pada butir ke- i hingga ke- n .

Estimasi parameter butir dalam teori respon butir sangat penting untuk dilakukan. Estimasi ini dilakukan untuk mengetahui apakah suatu butir soal cocok dengan model atau tidak. Semakin banyak butir yang cocok dengan suatu model, akan semakin baik analisis yang dilakukan dengan menggunakan model tersebut.

Estimasi parameter kemampuan dalam teori respon butir juga sangat penting dilakukan untuk mengestimasi tingkat kemampuan masing-masing siswa. Tingkat kemampuan antar siswa pasti berbeda pada setiap butir soal suatu perangkat tes. Seorang pendidik sudah seharusnya mengetahui tingkat kemampuan masing-masing siswanya agar dapat menerapkan metode yang cocok dan materi yang menyesuaikan agar dapat diterima oleh peserta didik. Penelitian ini menggunakan metode EAP (*Expected a Posteriori*) untuk menghitung estimasi parameter kemampuan (θ). Selain metode EAP, dapat juga menggunakan metode lain seperti *Maximum Likelihood Estimation* (MLE) dan *Maximum a Posteriori* (MAP).

HASIL DAN PEMBAHASAN

Hasil

Data penelitian merupakan hasil dari pengerjaan siswa terhadap suatu perangkat tes yakni penilaian harian mata pelajaran Matematika di sekolah dasar. responden berjumlah 329 siswa dengan data penelitian dikotomi, yakni nilai 1 untuk jawaban benar dan nilai 0 untuk jawaban siswa yang salah. Hasil pengukuran kemampuan tes penilaian harian diestimasi dengan menggunakan uji kecocokan model dengan model Rasch, model 1PL, model 2PL, model 3PL, dan model 4PL. Data juga akan diestimasi berdasarkan grafik ICC (*Item Characteristic Curve*).

Besarnya tingkat kesukaran soal berkisar antara 0 sampai 1. Apabila proporsi menjawab benar (p) mendekati 0 maka soal akan semakin tergolong sulit. Begitupun apabila nilai p mendekati 1 maka soal akan semakin tergolong mudah. Menurut perhitungan nilai biserial butir soal, tidak didapati nilai biserial negatif pada semua butir soal sehingga seluruh butir soal akan digunakan dalam penelitian ini. Hal tersebut juga berarti penelitian ini menggunakan seluruh data yang tersedia. Namun, pada butir soal nomor 7 terlihat nilai biserial yang sangat kecil. Hal tersebut dapat menunjukkan bahwa terdapat kemungkinan soal tersebut mungkin tidak dapat terlalu mengukur kemampuan responden apakah tinggi atau rendah. Butir dengan nilai biserial yang rendah mungkin juga tidak memberikan informasi yang signifikan dalam memberikan informasi terkait kemampuan siswa.

Data penelitian selanjutnya diestimasi dengan menggunakan model Rasch, 1PL, 2PL, 3PL, dan 4PL. Hasil perhitungan masing-masing model akan dibandingkan berdasarkan jumlah butir yang cocok dengan model tersebut. Model Rasch dan 1PL hanya mengestimasi satu karakteristik butir soal saja yakni tingkat kesulitan (*difficulty*). Model 2PL mengestimasi dua karakteristik butir soal yakni tingkat kesulitan (*difficulty*), daya beda (*discrimination*) butir soal. Model 3 PL mengestimasi tiga karakteristik butir soal yakni tingkat kesulitan (*difficulty*), daya beda (*discrimination*), dan tebakan semu (*pseudo guessing*). Sedangkan model 4 PL mengestimasi empat karakteristik butir soal yakni tingkat kesulitan (*difficulty*), daya beda (*discrimination*), tebakan semu (*pseudo guessing*), dan kecerobohan (*carelessness*).

Hasil uji kecocokan model Rasch menunjukkan bahwa terdapat enam butir soal yang cocok dan 14 butir soal tidak cocok. Dalam penelitian ini akan dibandingkan hasil perhitungan antara model Rasch, 1PL, 2PL, 3PL, dan 4PL untuk memilih model terbaik. Hasil kecocokan

model 1PL terlihat bahwa hanya terdapat 6 butir soal yang cocok ketika menggunakan model 1PL. Hal ini juga berarti terdapat 14 butir lainnya yang tidak cocok ketika menggunakan model 1PL. Pada perhitungan dengan model 2PL, didapatkan 17 butir soal yang cocok dan 3 butir soal yang tidak cocok. Model 3PL menunjukkan adanya 16 butir soal yang cocok dan empat butir soal lainnya tidak cocok. Sedangkan menggunakan model 4PL didapatkan 16 butir yang cocok, 3 butir tidak cocok, dan 1 butir NA. Terdapat satu butir soal dengan keterangan NA dikarenakan adanya kesamaan nilai data. Pada butir soal nomor satu seluruh siswa mayoritas menjawab benar. Hal tersebut mengakibatkan tidak banyaknya variasi jawaban siswa sehingga nilai kecocokan model tidak terdeteksi dan pada penelitian ini diasumsikan cocok dengan model 4PL. Setelah dilakukan uji kecocokan model pada masing-masing model maka didapatkan hasil seperti pada Tabel 1.

Tabel 1. Perbandingan hasil uji kecocokan model

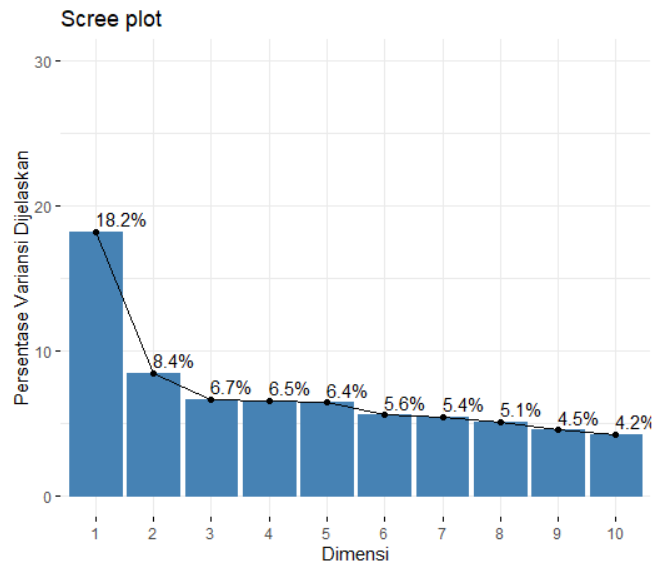
	Model Rasch	Model 1PL	Model 2PL	Model 3PL	Model 4PL
Cocok	6	6	17	16	17
Tidak Cocok	14	14	3	4	3

Berdasarkan Tabel 1 diperoleh model dengan jumlah cocok paling banyak terdapat pada model 2PL dan 4PL. Namun karena pada model 4PL terdapat 1 butir dengan nilai NA dan kemungkinan dapat mengganggu proses perhitungan, maka penelitian ini menggunakan model 2PL sebagai model yang paling cocok.

Tabel 2. Hasil uji kecocokan model 2PL

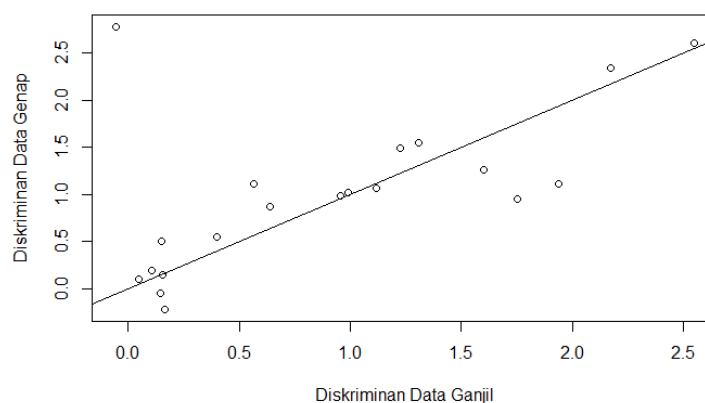
	Butir	Chi-square	p-value	Keputusan
1	b1	1,28	0,87	Cocok
2	b2	6,34	0,79	Cocok
3	b3	19,62	0,05	Cocok
4	b4	12,29	0,20	Cocok
5	b5	7,42	0,69	Cocok
6	b6	21,87	0,02	Tidak Cocok
7	b7	20,48	0,08	Cocok
8	b8	7,95	0,85	Cocok
9	b9	15,30	0,12	Cocok
10	b10	18,32	0,11	Cocok
11	b11	10,72	0,38	Cocok
12	b12	12,63	0,13	Cocok
13	b13	11,14	0,60	Cocok
14	b14	15,11	0,30	Cocok
15	b15	31,73	0,00	Tidak Cocok
16	b16	10,02	0,26	Cocok
17	b17	11,15	0,52	Cocok
18	b18	16,59	0,08	Cocok
19	b19	11,50	0,49	Cocok
20	b20	24,52	0,01	Tidak Cocok

Asumsi-asumsi yang harus terpenuhi dalam teori respon butir di antaranya adalah asumsi unidimensi, asumsi invariansi parameter, dan asumsi independensi lokal. Unidimensi berarti setiap butir tes hanya mengukur satu kemampuan saja (Retnawati, 2014). Perhitungan uji asumsi unidimensi menggunakan *function PCA()* pada *software* RStudio dengan menggunakan *eigenvalue* pada perhitungannya.



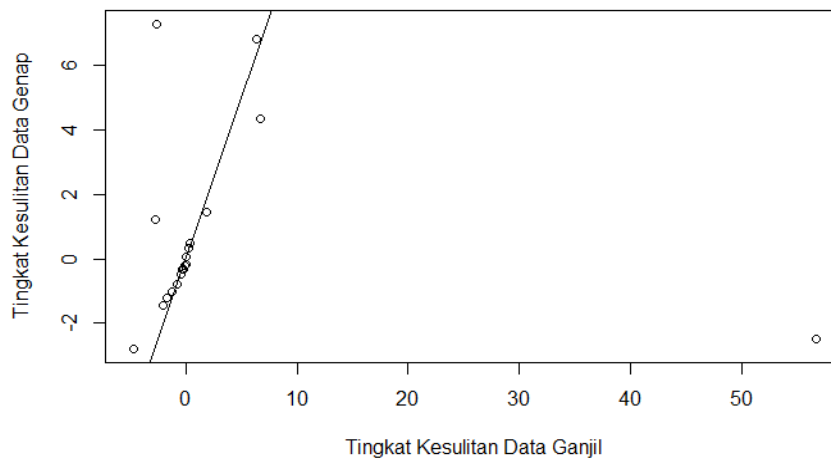
Gambar 1. Scree plot data

Scree plot data pada Gambar 1 menunjukkan adanya dominasi satu data dengan data lainnya serta terdapat satu titik elbow di mana terdapat penurunan yang semakin sedikit pada data tersebut. Hal ini menunjukkan bahwa data dapat dikatakan unidimensi atau asumsi unidimensi terpenuhi.



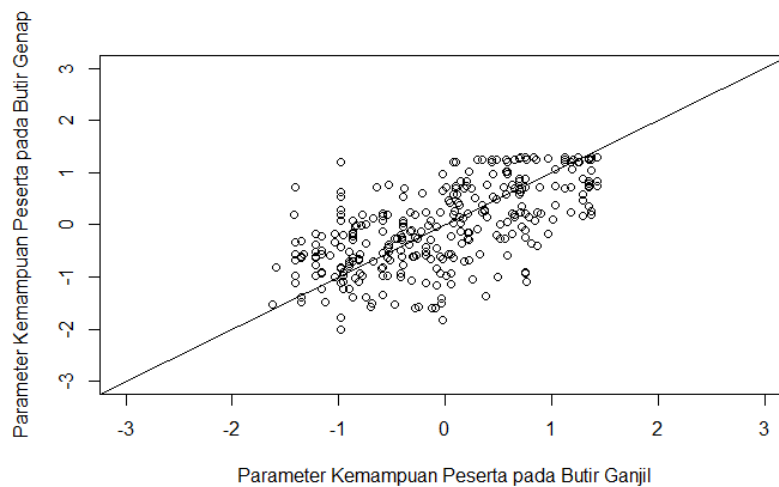
Gambar 2. Invariansi parameter butir daya beda

Gambar 2 tentang uji asumsi invariansi parameter butir daya beda menunjukkan adanya titik-titik yang menyebar mengikuti garis lurus. Hal tersebut menunjukkan bahwa asumsi invariansi parameter butir daya beda pada data hasil penilaian harian materi pecahan pada siswa sekolah dasar terpenuhi.



Gambar 3. Plot uji asumsi invariansi parameter tingkat kesulitan

Plot uji asumsi invariansi parameter butir tingkat kesulitan di atas menunjukkan adanya titik-titik yang menyebar mengikuti garis lurus. Hal tersebut menunjukkan bahwa asumsi invariansi parameter butir tingkat kesulitan pada data hasil penilaian harian materi pecahan pada siswa sekolah dasar terpenuhi.



Gambar 4. Plot uji asumsi invariansi parameter kemampuan

Plot uji asumsi invariansi parameter butir kemampuan di atas menunjukkan adanya titik-titik yang menyebar mengikuti garis lurus. Hal tersebut menunjukkan bahwa asumsi invariansi parameter butir kemampuan pada data hasil penilaian harian materi pecahan pada siswa sekolah dasar terpenuhi.

Asumsi independensi lokal pada data hasil penilaian harian materi pecahan di sekolah dasar dapat dikatakan terpenuhi. Hal tersebut dikarenakan asumsi unidimensi sudah terpenuhi. Perhitungan uji asumsi independensi lokal pada penelitian ini menggunakan bantuan RStudio.

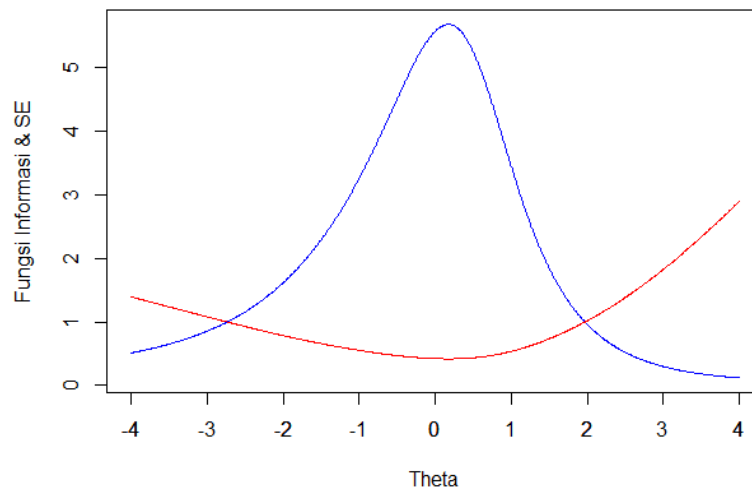
Berdasarkan model pada teori respon butir yang digunakan, dalam penelitian ini menggunakan model 2PL sebagai model paling cocok. Estimasi karakteristik butir soal pada perangkat tes menurut model 2PL adalah seperti pada Tabel 3.

Tabel 3. Karakteristik butir berdasarkan model terbaik

	<i>a</i>	<i>b</i>	Keterangan
b1	0,59645976	-6,26115870	Kurang baik
b2	1,34383849	0,05909932	Baik
b3	1,04916396	-0,31416432	Baik
b4	1,45436530	-0,27416115	Baik
b5	0,97307047	-1,44801470	Baik
b6	1,36962773	-0,46870358	Baik
b7	-0,02500673	14,50168495	Kurang baik
b8	0,04738583	-7,65536559	Kurang baik
b9	1,29159779	-0,03925200	Baik
b10	0,46411608	-1,71421142	Baik
b11	1,42365430	0,27203826	Baik
b12	0,82578814	-3,46775563	Kurang baik
b13	0,15327799	6,60817956	Kurang baik
b14	0,14709142	5,36955256	Kurang baik
b15	2,53976017	0,45457455	Kurang baik
b16	2,22694230	0,03980217	Kurang baik
b17	0,07521680	1,53932275	Baik
b18	1,10273416	-1,10092511	Baik
b19	0,31944347	-0,17634129	Baik
b20	0,73126977	-0,79209826	Baik

Menurut Hambleton dan Swaminathan (1985), parameter daya beda (*a*) dikatakan baik apabila berada dalam interval $0 < a < 2$, sedangkan parameter kesukaran butir (*b*) dikatakan baik apabila nilainya berada dalam interval $-2 < b < 2$. Apabila nilai b_i mendekati 2, tingkat kesulitan soal semakin tinggi. Begitu pula apabila nilai b_i mendekati -2, tingkat kesulitan soal semakin rendah. Maka dari itu, didapatkan hasil seperti pada tabel karakteristik butir soal tersebut.

Fungsi informasi pada level butir soal menyatakan suatu kejelasan atau kekuatan butir soal dalam menjelaskan kemampuan siswa/responden atau yang dikenal sebagai *latent trait* yang diukur dengan pengukuran tes (Myszkowski, 2019). Sedangkan menurut Mulvia dkk. (2021), fungsi informasi merupakan suatu fungsi yang memberikan estimasi mengenai kemampuan responden dari model pada teori respon butir. *Standard Error of Measurement* (SEM) atau yang dikenal juga sebagai kesalahan pengukuran mengukur nilai kesalahan dalam suatu pengukuran.



Gambar 4. Plot fungsi informasi dan SEM

Plot fungsi informasi dan *standard error of measurement* (SEM) di atas menunjukkan bahwa nilai fungsi informasi maksimum dari penilaian harian mata pelajaran Matematika di sekolah dasar dengan 20 butir soal adalah 5,7 pada kemampuan siswa (θ) sekitar 0,2 dan SEM sebesar 0,4. Nilai fungsi Informasi akan lebih tinggi dari SEM ketika θ berada di antara -2,4 dan 1,8. Hal ini berarti perangkat tes prestasi yang digunakan dalam penelitian ini dapat memberikan informasi mengenai kemampuan peserta tes pada kategori sangat rendah hingga tinggi.

Pembahasan

Penelitian ini bertujuan untuk menganalisis karakteristik butir soal pada perangkat tes prestasi dan kemampuan siswa di sekolah dasar. Penelitian ini menggunakan perangkat tes prestasi berupa soal penilaian harian mata pelajaran Matematika dengan topik pecahan yang diberikan kepada siswa kelas 4 sekolah dasar. Data yang digunakan berupa data dikotomi hasil pengerjaan siswa pada soal pilihan ganda dengan nilai 0 untuk jawaban salah dan nilai 1 untuk jawaban yang benar. Total peserta yang mengikuti tes tersebut sebanyak 329 siswa yang berasal dari 12 sekolah berbeda.

Analisis tingkat kesukaran butir dalam suatu instansi pendidikan bertujuan untuk mengetahui tingkat kesulitan butir-butir soal apakah termasuk dalam kategori soal mudah, soal sedang, atau soal yang sulit (Bagiyono, 2017). Apabila $p > 0,7$ maka tergolong soal mudah, apabila nilai p berada pada $0,3 \leq p \leq 0,7$ maka tergolong soal sedang, sedangkan apabila $p < 0,3$ maka tergolong soal yang sulit. Kategori ini selaras dengan pendapat Allen dan Yen (1989) yang menyatakan bahwa sebaiknya indeks kesukaran butir berada pada interval 0,3 – 0,7 agar informasi mengenai kemampuan siswa dapat diperoleh secara maksimal. Hasil perhitungan menunjukkan sebanyak 3 butir soal tergolong dalam kategori soal mudah yakni butir soal b1, b5, dan b12. Terdapat 1 butir tergolong dalam soal sulit yakni b13. Sedangkan 16 butir soal lainnya tergolong dalam soal sedang.

Daya beda dapat dilihat dari besarnya koefisien korelasi biserial maupun koefisien korelasi *point* biserial. Nilai biserial merujuk pada tingkat daya beda suatu butir soal yang diujikan kepada siswa. Langkah pengujian dalam penelitian ini dibantu dengan *software* RStudio dalam *package* CTT. Hasil perhitungan pada penelitian ini menunjukkan bahwa 20 butir soal dapat dipakai dalam penelitian karena tidak ada butir soal dengan nilai biserial negatif. Namun, pada butir soal b7 nilai biserial sangatlah rendah yakni 0,097. Hal tersebut

menunjukkan bahwa terdapat kemungkinan soal tersebut mungkin tidak dapat terlalu mengukur kemampuan responden apakah tinggi atau rendah. Butir dengan nilai biserial yang rendah mungkin juga tidak memberikan informasi yang signifikan dalam memberikan informasi terkait kemampuan siswa. Hal ini seperti yang disampaikan oleh Amelia & Kriswanto (2017) bahwa item dengan nilai koefisien biserial negatif tidak dimasukkan ke dalam data penelitian karena dapat mengganggu proses analisis.

Data hasil pengerjaan siswa diuji berdasarkan teori respon butir dengan terlebih dahulu menentukan model yang paling cocok dengan data. Berdasarkan uji kecocokan model teori respon butir, diperoleh model 2 parameter logistik (2PL) sebagai model yang dianggap paling cocok untuk digunakan dalam penelitian. Terdapat 17 butir soal cocok dan 3 butir soal tidak cocok dengan model 2PL.

Teori respon butir memiliki beberapa asumsi yang harus dipenuhi diantaranya asumsi unidimensi, asumsi invariansi parameter, dan asumsi independensi lokal. Hasil dari penelitian ini menerangkan bahwa adanya satu dominasi data terhadap data lain yang menunjukkan bahwa setiap butir dalam perangkat tes tersebut hanya mengukur satu kemampuan saja. Hal ini menunjukkan bahwa asumsi unidimensi pada penelitian ini dapat terpenuhi. Pada pengujian asumsi invariansi parameter dilakukan uji invariansi parameter butir daya beda (a), uji invariansi parameter butir tingkat kesulitan (b), serta uji invariansi parameter kemampuan (θ). Seluruh uji invariansi parameter yang dilakukan menghasilkan plot yang menunjukkan adanya titik-titik yang menyebar mengikuti garis lurus. Hal tersebut berarti seluruh uji asumsi invariansi parameter pada penelitian ini dapat terpenuhi. Asumsi independensi lokal pada penelitian ini terpenuhi ketika asumsi unidimensi dapat terpenuhi.

Setelah semua pengujian asumsi dapat terpenuhi, dapat dilanjutkan pengujian untuk melihat bagaimana karakteristik perangkat tes prestasi dan kemampuan siswa berdasarkan model 2 parameter logistik teori respon butir. Pada pengujian ini didapati 8 butir soal yang kurang baik dan 12 butir soal yang dapat dikatakan baik. Adanya butir-butir soal yang kurang baik menunjukkan bahwa analisis mengenai karakteristik butir soal pada penilaian harian penting dilakukan. Hal tersebut juga sebagai standar pengukuran kemampuan siswa dalam menghadapi masalah pada suatu topik pembelajaran.

Nilai daya beda pada penelitian ini mayoritas sudah masuk dalam kategori baik karena berada dalam interval $0 < a < 2$. Namun, pada butir soal b7, b15, dan b16, tingkat daya beda masih belum baik dikarenakan nilai daya beda tidak masuk dalam interval tersebut. Butir soal nomor b15 dan b16 menunjukkan tingkat daya beda yang kurang baik karena $a > 2$, sedangkan butir soal nomor b7 menunjukkan tingkat daya beda yang kurang baik karena nilai $a < 0$.

Hasil perhitungan tingkat kesulitan pada penelitian ini menunjukkan terdapat enam butir soal yang kurang baik karena nilai tingkat kesulitan tidak berada dalam interval $-2 < b < 2$. Butir soal yang tingkat kesulitannya kurang baik yakni butir soal nomor b1, b7, b8, b12, b13, dan b14. Pada butir soal nomor b1, b8, dan b12 terlihat nilai tingkat kesulitan yang cukup rendah dan menunjukkan bahwa soal pada butir-butir tersebut terlalu mudah dan perlu diperbaiki. Pada butir soal nomor b7, b13, dan b14 terdapat nilai tingkat kesulitan yang cukup tinggi yang menunjukkan soal terlalu sulit untuk diujikan ke siswa dan perlu diperbaiki.

Penelitian ini menggunakan model 2PL yang mengestimasi karakteristik daya beda (a) dan tingkat kesulitan (b) butir soal sehingga akan dianalisis karakteristik soal menurut dua parameter tersebut. Butir b1, b8, dan b12 termasuk dalam kategori soal yang kurang baik dikarenakan nilai $b < -2$. Butir soal b13 dan b14 termasuk dalam kategori soal yang kurang baik dikarenakan nilai $b > 2$. Butir soal b15 dan b16 termasuk dalam kategori soal yang kurang baik dikarenakan nilai $a > 2$, sedangkan butir soal b7 termasuk dalam kategori soal yang kurang baik dikarenakan nilai $a < 0$ dan nilai $b > 2$. Selain pada butir-butir soal tersebut, butir

soal b2, b3, b4, b5, b6, b9, b10, b11, b17, b18, b19, dan b20 sudah termasuk dalam kategori soal yang baik. Hal tersebut dikarenakan pada butir-butir soal tersebut nilai daya beda berada dalam interval $0 < a < 2$ dan nilai tingkat kesulitan butir berada dalam interval $-2 < b < 2$. Hal ini menunjukkan walaupun nilai daya beda (a) sudah baik tapi nilai tingkat kesulitan (b) masih kurang baik, maka butir soal tersebut tetap masih kurang baik. Hal tersebut seperti yang telah disampaikan oleh Santoso (2018) dalam penelitiannya yang menyimpulkan sebab banyaknya butir soal yang kurang baik.

Analisis terhadap suatu perangkat tes sangat penting untuk dilakukan agar perangkat tes tersebut benar-benar dapat mengukur kemampuan siswa dan kualitas soal tersebut. Seperti yang disampaikan oleh Alfarisa & Purnama (2019) bahwa penilaian yang baik itu penting bagi setiap guru dan diharapkan masing-masing guru menerima pelatihan mengenai analisis butir soal. Teori respon butir akan terus berkembang sebagai metode analisis perangkat tes prestasi (Andayani, dkk. 2019). Hal ini dikarenakan teori respon butir dapat lebih unggul dari teori tes klasik karena asumsi-asumsi teori respon butir dapat menutupi kelemahan dalam teori tes klasik.

Penelitian ini memiliki fungsi informasi maksimum dari penilaian harian mata pelajaran Matematika di sekolah dasar dengan 20 butir soal adalah 5,7 pada kemampuan siswa (θ) sekitar 0,2 dan SEM sebesar 0,4. Nilai fungsi Informasi akan lebih tinggi dari SEM ketika θ berada di antara -2,4 dan 1,8. Hal ini berarti perangkat tes prestasi yang digunakan dalam penelitian ini dapat memberikan informasi mengenai kemampuan peserta tes pada kategori sangat rendah hingga tinggi. Hal ini juga serupa dengan penelitian yang telah dilakukan oleh Harsana (2024) pada interval antara titik-titik potong pada kurva fungsi informasi terdapat nilai SEM yang lebih besar sehingga perangkat tes baik digunakan pada tingkat kemampuan sesuai dengan interval tersebut. Hasil ini berbeda dengan penelitian yang telah dilakukan oleh Alifa, dkk. (2018) dan menghasilkan nilai fungsi informasi yang cukup tinggi dan perangkat tes baik untuk siswa dengan tingkat kemampuan yang tinggi saja.

SIMPULAN

Berdasarkan hasil penelitian terkait karakteristik perangkat tes prestasi dan kemampuan siswa di sekolah dasar berdasarkan teori respon butir, dapat disimpulkan bahwa model yang paling cocok digunakan dalam penelitian ini adalah model 2 parameter logistik (model 2PL). Karakteristik yang diestimasi oleh model 2PL yakni tingkat daya beda (a) dan tingkat kesulitan butir soal (b). Hasil pengujian menunjukkan bahwa dengan model 2PL berdasarkan teori respon butir didapatkan sebanyak 8 butir soal yang kurang baik dan 12 butir soal yang sudah baik. Hal tersebut juga menunjukkan pentingnya pengujian perangkat tes pada penilaian hasil belajar siswa agar kemampuan siswa dapat terukur secara maksimal.

DAFTAR PUSTAKA

- Alfarisa, F., & Purnama, D.N. (2019). Analisis butir soal ulangan akhir semester mata pelajaran ekonomi sma menggunakan rasch model. *Jurnal Pendidikan Ekonomi*, 11(2).
- 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*.

- Allen, M.J., & Yen, W.M. (1989). *Introduction to measurement theory*. California: Brokes/Cole Publishing Company.
- Amelia, R.N., & Kriswanto, (2017). Implementasi item response theory sebagai basis analisis kualitas butir soal dan kemampuan kimia siswa kota yogyakarta. *JKPK (Jurnal Kimia dan Pendidikan Kimia)*, 2(1).
- Andayani, A., Purwanto, & Ramalis, T.R. (2019). Kajian implementasi teori respon butir dalam menganalisis instrumen tes materi fisika. *Prosiding Seminar Nasional Fisika 5.0*
- Bagiyono, (2017). Analisis tingkat kesukaran dan daya pembeda butir soal ujian pelatihan radiografi tingkat 1. *Widyanuklida*, 16(1).
- Hambleton, R.K., (1993). *Principles and selected application of item response theory*. In Linn. & Robert, L. (Eds.). *Educational measurement*. Third Edition. Phoenix: American Council on Education, Series on Higher Education Oryx Press.
- Hambleton, R.K., Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer.
- Hambleton, R.K., Swaminathan, H., dan Rogers, J. (1991). *Fundamentals of item response theory*. London: Sage Publications.
- Harsana, F.N. (2024). *Perbandingan analisis karakteristik butir soal literasi membaca dengan teori respon butir politomi*. [Skripsi]. Universitas Negeri Yogyakarta.
- Lord, F.M. (1990). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates.
- Mulvia, R., Ramalis, T.R., & Evendi, R., (2021). Mendeteksi keajegan butir tes dengan fungsi informasi. *Jurnal Pendidikan Indonesia*. 2(1).
- 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.
- Retnawati, H. (2014). *Teori respons butir dan penerapannya: Untuk peneliti, praktisi pengukuran dan pengujian, mahasiswa pascasarjana*. Yogyakarta: Nuha Medika.
- Santoso, A. (2018). Karakteristik butir tes pengantar statistika sosial berdasarkan teori respon butir. *Jurnal Pendidikan Matematika dan Sains*, 6 (2), 158-168.
- Sudjana, N. (2010). *Penilaian hasil proses belajar mengajar*. Bandung: Remaja Rosakarya.
- Sugiyono. (2010). *Metodologi penelitian kuantitatif, kualitatif, dan r&d*. Bandung: Alfabeta.