JURNAL STATISTIKA DAN SAINS DATA
Volume 2 Nomor 1, Oktober, 2024, 65-80
https://journal.student.uny.ac.id/index.php/jssd

PERBANDINGAN KECOCOKAN MODEL ANALISIS BUTIR DATA DIKOTOMI DENGAN TEORI RESPON BUTIR

Novi Sheila Wardhani*, Universitas Negeri Yogyakarta
Heri Retnawati, Universitas Negeri Yogyakarta
*e-mail: novisheila.2019@student.uny.ac.id

Abstrak. Penelitian ini merupakan penelitian kuantitatif untuk mengetahui karakteristik butir soal suatu perangkat tes dan mendeskripsikan kemampuan hasil ujian tulis Bahasa Inggris mahasiswa Indonesia. Data yang digunakan adalah data sekunder sebanyak 1500 respons peserta tes terhadap soal ujian tulis yang menempuh bahasa Inggris manajemen pada ujian akhir semester di sebuah universitas negeri tahun 2022 se-Indonesia. Teknik analisis data dengan teori respons butir dikotomi meliputi kecocokan model *Rasch*, 1PL, 2PL, 3PL, dan 4PL yang dilakukan dengan metode *Yen's Q1* serta memperhatikan nilai *Akaike Information Criterion (AIC)* dan *Bayesian Information Criterion* (BIC). Hasil penelitian ini menunjukkan bahwa: (1) Kecocokan model *Rasch* sebanyak 6 butir; (2) kecocokan model 1PL sebanyak 6 butir; (3) kecocokan model 2PL; (4) sebanyak 11 butir; (5) kecocokan model 3PL sebanyak 19 butir; (6) kecocokan model 4PL sebanyak 19 butir. Selain itu, pada model 4PL menunjukkan nilai terendah untuk AIC dan BIC, dengan nilai AIC sebesar 37967,02 dan nilai BIC sebesar 38498,34. Nilai fungsi informasi yang maksimum sebesar 18,51 pada kemampuan (θ) sebesar 1,07 dengan tingkat kesalahan pengukuran baku (SEM) sebesar 1. Selain itu, nilai fungsi informasi akan lebih tinggi daripada SEM ketika kemampuan θ berkisar -2,8 hingga +1,8.

Kata kunci: Teori Respons Butir, Tes Kemampuan Ujian Tulis Bahasa Inggris.

Abstract. *This study is a quantitative research to determine the item characteristics of a test device and describe the ability of Indonesian students' English written test results. The data used is secondary data as many as 1500 test takers' responses to written exam questions who took English management at the end of semester exams at a state university in 2022 throughout Indonesia. Data analysis techniques with dichotomous item response theory include Rasch model fit, 1PL, 2PL, 3PL, and 4PL conducted by Yen's Q1 method and pay attention to the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. The results of this study show that: (1) Rasch model fit is 6 items; (2) 1PL model fit is 6 items; (3) 2PL model fit; (4) 11 items; (5) 3PL model fit is 19 items; (6) 4PL model fit is 19 items. In addition, the 4PL model shows the lowest value for AIC and BIC, with an AIC value of 37967.02 and a BIC value of 38498.34. The maximum value of the information function was 18.51 at an ability (θ) of 1.07 with a standardized measurement error (SEM) of 1. In addition, the value of the information function will be higher than the SEM when the ability θ ranges from -2.8 to +1.8.*

Keywords Item Response Theory, English Language Writing Examination Proficiency Test.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

PENDAHULUAN

Pendidikan merupakan proses pembelajaran untuk peserta didik mengembangkan potensi dirinya agar memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa, dan negara. Untuk melihat perkembangan potensi peserta didik, maka diakhir semester dilakukan penilaian. Penilaian sangat penting dalam pengembangan kinerja siswa ketika menggunakan instrumen yang baik yang dirancang di kelas (Kasali et. al., 2022). Salah satu bentuk penilaian hasil belajar peserta didik adalah Ujian Akhir Semester (UAS). UAS adalah kegiatan yang dilakukan oleh pendidik untuk mengukur pencapaian peserta didik di akhir semester. Dalam dunia pendidikan, sebagian besar ujian akhir semester berupa tes pilihan ganda dengan cara penskoran dikotomi. Penskoran dikotomi berarti ketika peserta didik menjawab butir dengan benar maka diberikan skor 1, sedangkan jika salah diberikan skor 0. Pengumpulan respon peserta tes sering dilakukan dengan penskoran dikotomi (Anisa, 2013) yaitu peserta tes menjawab butir soal tes berbentuk pilihan ganda dengan benar diberi skor 1, dan salah diberi skor 0, dengan adanya respons peserta yang telah dikumpulkan, maka kemampuan peserta tes dapat diukur dengan mengestimasi parameter kemampuan yang kemudian dilakukan uji kecocokan model.

Teori respons butir menjadi alternatif yang dapat digunakan untuk menganalisis suatu tes (Sudaryono, 2011). Jika probabilitas dinyatakan dengan P, kemampuan dengan (θ), pembeda butir (a), tingkat kesulitan (b), dan tebakan semu (*pseudo guessing, c)*, maka hubungan kelima besaran tersebut dinyatakan dengan persamaan (Hambleton, Swaminathan, & Rogers, 1991). Persamaan inilah yang dinyatakan sebagai model yang memuat 4 parameter yang didefinisikan sebagai model 4PL (model logistik empat parameter). Model Rasch dan 1PL hanya mengestimasi satu karakteristik butir yakni tingkat kesulitan (b) saja. Model 2PL mengestimasi dua karakteristik butir yakni daya beda (a) dan tingkat kesulitan (b). Model 3PL memiliki tiga karakteristik butir yang akan diestimasi yakni daya beda (a), tingkat kesulitan (b), dan tebakan semu (c). Model 4PL memiliki empat karakteristik butir yakni daya beda, tingkat kesulitan, tebakan semu dan kecerobohan (Fatkhudin et.al., 2014). Hasil perhitungan masing-masing model akan dibandingkan berdasarkan jumlah banyaknya butir yang cocok dengan model tersebut.

METODE

Dalam penelitian ini, data yang digunakan merupakan data dokumenter berupa hasil ujian tulis bahasa Inggris manajemen pada ujian akhir semester se-Indonesia pada tahun 2022. Data diperoleh dari pihak perguruan tinggi yang menyelenggarakan ujian tulis bahasa Inggris niaga pada ujian akhir semester dari 4.836 diambil acak sebanyak 1.500.

Untuk tahapan data analisis data yang dilakukan adalah sebagai berikut:

1. Mempersiapkan data perangkat tes kemampuan *reading* bahasa Inggris.
2. Menentukan model terbaik dengan uji kecocokan model antara model *Rasch*, 1PL, 2PL, 3PL, dan 4PL. Kemudian memilih butir paling banyak yang cocok dan melihat plot kurva karakteristik butir menggunakan fungsi mirt, coef, fscores, itemfit pada *package* mirt (Chalmers, 2012).
3. Uji asumsi teori respons butir yaitu unidimensi dengan menggunakan analisis nilai eigen dari matriks korelasi antarbutir, independensi lokal dapat terpenuhi apabila asumsi unidimensi terpenuhi dan invariansi parameter dilakukan dengan membagi

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

data menjadi dua kelompok yaitu ganjil dan genap kemudian diestimasi butir model paling cocok menghasilkan parameter daya beda (*a*) dan tingat kesulitan (*b*) kemudian dibuat plot dan pada invariansi parameter kemampuan dilakukan dengan membagi jumlah butir hasil jawaban menjadi dua bagian yaitu butir hasil jawaban peserta genap dan hasil jawaban peserta ganjil dengan bantuan R menggunakan fungsi PCA, get_eigenvalue, dan fviz_eig pada *package* factoextra (Kassambra, 2016) serta factorMineR (Le, Josse, & Husson, 2008).

4. Mengestimasi parameter butir (karakteristik perangkat tes) yang menghasilkan parameter daya beda (*a*) dan tingkat kesulitan (*b*) dan mengestimasi parameter kemampuan yang menghasilkan nilai kemampuan dari model terbaik.
5. Menentukan model yang paling fit dengan melihat nilai terkecil dari hasil AIC dan BIC.
6. Menentukan model yang paling cocok yang digunakan pada data penelitian.

Penentuan model yang paling cocok dilakukan dengan menggunakan analisis menggunakan setiap model yang digunakan yaitu model *Rasch*, 1PL, 2PL, 3PL, dan 4PL. Setiap model terlebih dahulu menentukan estimasi tingkat kesulitan, nilai diskriminan, dan estimasi kemampuan dari peserta tes. Kemudian dilakukan analisis dengan model dan dilakukan pencarian khi-kuadrat *Yen's Q1* yang digunakan untuk penentuan kecocokan butir. Dari hasil kecocokan butir tersebut dilakukan perbandingan model dan dilihat model yang paling cocok.

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* memprediksi probabilitas jawaban benar dengan menggunakan satu parameter saja yakni parameter tingkat kesulitan butir (*b*). Persamaan kurva karakteristik item pada model *Rasch* ditunjukkan sebagai berikut (Paek & Cole, 2020).

$$P\left(X_{ij}=1|\theta_j\right)=\frac{exp^{(\theta_j-b_i)}}{1+exp^{(\theta_j-b_i)`}} \quad (1)$$

dengan:

i	:	1,2, …, m
j	:	1,2, …, n
m	:	banyaknya butir tes
n	:	banyaknya peserta tes
$P\left(X_{ij}=1\right)$	:	probabilitas jawaban peserta ke-j menjawab benar pada butir ke-i
b_i	:	parameter kesulitan butir ke-i
e	:	konstanta dengan nilai berkisar 2.718
θ_j	:	parameter kemampuan peserta ke- j

2. Model 1PL

Model logistik satu parameter merupakan probabilitas peserta tes menjawab benar butir soal didefinisikan oleh satu karakteristik butir yaitu tingkat kesulitan butir. Model ini secara matematis sama seperti model *Rasch*, namun perbedaanya terletak pada nilai daya beda (*a*) untuk model *Rasch.* Persamaan kurva karakteristik item menurut Paek & Cole (2020) pada model 1PL dinyatakan sebagai berikut.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

$$P(X_{ij} = 1|\theta_j) = \frac{exp\,[a(\theta_j - b_i)]}{1 + exp\,exp\,[a(\theta_j - b_i)]`} \quad (2)$$

dengan:

i	:	$1,2,\ldots,m$
j	:	$1,2,\ldots,n$
m	:	banyaknya butir tes
n	:	banyaknya peserta tes
$P(X_{ij} = 1)$	:	probabilitas jawaban peserta ke-j menjawab benar pada butir ke-i
b_i	:	parameter kesulitan butir ke-i
e	:	konstanta dengan nilai berkisar 2.718
θ_j	:	parameter kemampuan peserta ke- j
a_i	:	Parameter daya pembeda butir untuk butir ke-i

3. Model 2PL

Model logistik dua parameter (2PL) memprediksi jawaban benar dengan menggunakan dua parameter yakni parameter tingkat kesulitan butir (b) dan parameter daya beda butir (a). Semakin tinggi nilai daya beda butir, maka akan semakin baik butir tersebut untuk mendeteksi perbedaan halus kemampuan peserta tes. Menurut Paek & Cole (2020) bentuk persamaan pengukuran yang digunakan pada model 2 PL adalah sebagai berikut:

$$P(X_{ij} = 1|\theta_j) = \frac{exp\,[a_i(\theta_j - b_i)]}{1 + exp\,exp\,[a_i(\theta_j - b_i)]`} \quad (3)$$

dengan:

i	:	$1,2,\ldots,m$
j	:	$1,2,\ldots,n$
m	:	banyaknya butir tes
n	:	banyaknya peserta tes
$P(X_{ij} = 1)$	:	probabilitas jawaban peserta ke-j menjawab benar pada butir ke-i
b_i	:	parameter kesulitan butir ke-i
e	:	konstanta dengan nilai berkisar 2.718
θ_j	:	parameter kemampuan peserta ke- j
a_i	:	parameter daya pembeda butir untuk butir ke-i

4. Model 3PL

Menurut Paek & Cole (2020), bentuk persamaan pengukuran yang digunakan pada model 3PL adalah sebagai berikut.

$$P(X_{ij} = 1|\theta_j) = g_i + (1 - g_i)\frac{exp\,[a_i(\theta_j - b_i)]}{1 + exp\,exp\,[a_i(\theta_j - b_i)]`} \quad (4)$$

dengan:

i	:	$1,2,\ldots,m$
j	:	$1,2,\ldots,n$
m	:	banyaknya butir tes
n	:	banyaknya peserta tes
$P(X_{ij} = 1)$	:	probabilitas jawaban peserta ke-j menjawab benar pada butir ke-i

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

b_i : parameter kesulitan butir ke-i
e : konstanta dengan nilai berkisar 2.718
θ_j : parameter kemampuan peserta ke- j
a_i : parameter daya pembeda butir untuk butir ke-i
g_i : Parameter tebakan semu (*pseudo guessing*) butir ke-i

5. Model 4PL

Model 4PL mempunyai 5 parameter. Satu diantaranya adalah parameter untuk ciri peserta yaitu parameter kemampuan dan empat lainnya adalah parameter ciri butir yakni daya beda, tingkat kesulitan butir, parameter probabilitas menjawab benar secara kebetulan, dan parameter kecerobohan.

Secara matematis, model 4PL IRT adalah sebagai berikut

$$P_i(\theta) = c_i + (d_i - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad (5)$$

dengan:

$P_i(\theta)$: probabilitas peserta menjawab butir i
θ : kemampuan peserta (laten trait)
a_i : parameter daya pembeda butir ke$- i$
b_i : parameter tingkat kesulitan untuk butir i
c_i : asymptot bawah "guessing"
d_i : asymptot atas "carelessness"
e : bilangan natural $= 2{,}718$
D : faktor penskalaan $= 1{,}702$

Model terbaik pada penelitian ini didapat berdasarkan banyaknya butir yang cocok dengan model. Setelah mendapatkan model yang cocok, maka harus diuji asumsi berdasarkan teori respon butir. Penelitian 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 unidimensionalitas *(unidimensionality)* berarti bahwa hanya ada satu kemampuan yang diukur oleh seperangkat butir tes. Asumsi ini tidak dapat dipenuhi secara ketat karena adanya faktor-faktor kognitif, kepribadian, dan faktor-faktor administrasi tes, seperti kecemasan, motivasi, tendensi untuk menebak, dan semacamnya. Oleh karena itu, asumsi ini lebih diartikan bahwa hanya ada satu kemampuan yang dianggap paling dominan yang diukur oleh seperangkat butir di setiap tes.

2. Invariansi Parameter

Invariansi parameter merupakan karakteristik butir soal tidak tergantung pada distribusi parameter kemampuan peserta tes dan parameter yang menjadi ciri peserta tes tidak tergantung dari ciri butir soal (Retnawati, 2014). Invariansi parameter kemampuan (θ) dapat diselidiki dengan mengajukan dua perangkat tes atau lebih yang mempunyai tingkat kesulitan berbeda pada sekelompok peserta tes. Asumsi dapat dipenuhi apabila hasil estimasi kemampuan peserta tidak berbeda walaupun tingkat kesulitan dari tes dibedakan (Hambleton, 1991).

3. Independensi Lokal

Independensi lokal adalah independensi secara statistik. Apabila skor n butir dari peringkat tes dihasilkan oleh peserta tes (subyek) di dalam subpopulasi tersebut adalah

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

$u_1, u_2, \dots, u_n$, maka dari statistik diperoleh probabilitas skor peserta tes adalah hasil kali peluang menjawab semua butir soal tersebut.

$$P(u_1, u_2, \dots, u_n|\theta, \xi) = P(u_1|\theta, \xi)\, P(u_2|\theta, \xi) \dots P(u_n|\theta, \xi) = \prod_{i=1}^{n} P(u_i|\theta, \xi) \quad (6)$$

dengan:

i : $1,2, \dots, n$.
$P(U_i\,|\theta, \xi)$: probabilitas respon peserta tes yang mempunyai kemampuan θ.
u_i : jawaban peserta tes yang merupakan bilangan biner, 1 jika peserta tes menjawab dengan benar dan 0 jika peserta tes menjawab dengan salah

Asumsi ini berlaku jika peluang dari pola jawaban untuk setiap peserta tes adalah sama dengan hasil kali peluang jawaban peserta tes untuk setiap butir soal. Sebagai contoh, jika pola respon subyek pada 3 butir soal adalah (1,0,1), berarti bahwa $u_1 = 1$, $u_2 = 0$, dan $u_3 = 1$.

Asumsi independensi lokal terpenuhi jika:

$P(u_1 = 1, u_2 = 0, u_3 = 1|\theta, \xi) = P(u_1 = 1|\theta, \xi)\, P(u_2 = 0|\theta, \xi) P(u_3 = 1|\theta, \xi) = P_1 Q_2 Q_3$

dimana: $P_i = P(u_i = 1|\theta, \xi)$ dan $Q_i = 1 - P_i$

Asumsi independensi lokal disini berarti jawaban peserta tes terhadap butir soal yang berbeda dalam sebuah tes secara statistik adalah independen. Asumsi ini terpenuhi apabila jawaban subyek terhadap sebuah butir soal tidak mempengaruhi jumlah jawaban terhadap butir soal yang lain.

Tabel 1. Kriteria Model Teori Respons Butir (Hambleton et al, 1991)

Model	Kriteria Parameter			
	a_i	b_i	c_i	d_i
1PL	0 sampai +2	-	-	-
2PL	0 sampai +2	-2 sampai +2	-	-
3PL	0 sampai +2	-2 sampai +2	0 sampai $1/k$	-
4PL	0 sampai +2	-2 sampai +2	0 sampai $1/k$	mendekati 1

Keterangan $_i$ = indeks daya pembeda butir; b_i = indeks tingkat kesulitan butir**;** c_i = indeks *pseudo guessing* (menebak); d_i = indeks kecerobohan; k = banyaknya pilihan jawaban.

Estimasi parameter kemampuan dalam teori respon butir juga sangat penting dilakukan untuk mengestimasi tingkat kemampuan masing-masing mahasiswa. Tingkat kemampuan antar mahasiswa pasti berbeda pada setiap butir soal suatu perangkat tes.

.

HASIL DAN PEMBAHASAN

Hasil

Pada penelitian ini termasuk penelitian deskriptif. Data yang digunakan dalam penelitian ini adalah data hasil ujian tulis yang menempuh bahasa Inggris niaga pada ujian akhir semester se-Indonesia pada tahun 2022. Data yang digunakan dari 4.836 diambil acak sebanyak 1.500. Sebelum melakukan analisis dilakukan uji asumsi unidimensi dan independensi lokal pada data. Setelah itu melakukan estimasi dengan menggunakan model *Rasch*, 1PL, 2PL, 3PL,

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

dan 4PL. Hasil dari estimasi tersebut kemudian dilakukan perbandingan butir yang cocok dengan melihat khi-kuadrat lalu dipilih butir yang mempunyai kecocokan model paling banyak. Selain itu dilakukan perbandingan nilai AIC dan nilai BIC dimana model yang terbaik yang mempunyai nilai AIC dan BIC paling kecil. Selanjutnya dilakukan pengujian asumsi invariansi parameter untuk melihat uji asumsi tersebut terpenuhi. Sehingga diperoleh model yang paling cocok untuk digunakan.

Dalam penelitian ini akan dibandingkan hasil perhitungan antara model Rasch, 1PL, 2PL, 3PL, dan 4PL untuk memilih model terbaik. Hasil kecocokan model Rasch dan 1PL, terlihat bahwa hanya terdapat 6 butir soal yang cocok ketika menggunakan model Rasch dan 1PL. Hal ini juga berarti terdapat 19 butir lainnya yang tidak cocok ketika menggunakan model 1PL. Pada perhitungan dengan model 2PL, didapatkan 11 butir soal yang cocok dan 14 butir soal yang tidak cocok. Model 3PL menunjukkan adanya 19 butir soal yang cocok dan 9 butir soal lainnya tidak cocok. Sedangkan menggunakan model 4PL didapatkan 19 butir yang cocok, 6 butir tidak cocok. Sehingga nilai kecocokan model yang paling cocok dengan model 4PL. Dan nilai AIC sebesar 37967,02 dan nilai BIC sebesar 38498,34. Sehingga, model terbaik yang digunakan pada penelitian ini adalah model 4PL Setelah dilakukan uji kecocokan model pada masing-masing model maka didapatkan hasil seperti pada tabel model terbaik berikut.

Tabel 2. Ringkasan Hasil Uji Kecocokan pada Kelima Model

Keputusan	Rasch	1PL	2PL	3PL	4PL
Cocok	6	6	11	19	19
Tidak Cocok	19	19	14	6	6

Tabel 3. *Goodness of fit* AIC dan BIC

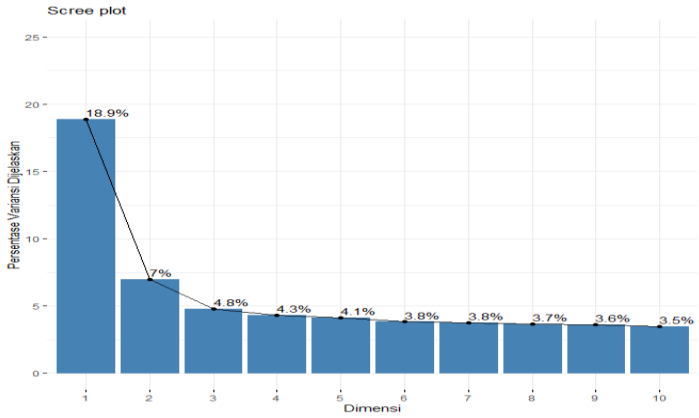
Model	AIC	BIC
Rasch	38891,69	39029,83
1PL	38891,69	39029,84
2PL	38378,28	38643,94
3PL	38082,75	38481,25
4PL	37967,02	38498,34

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Tabel 4. Kecocokan Model 4PL

item	S X2	df.S X2	RMSEA.S X2	p.S X2	Note
j1	16.25750286	16	0.003276651	0.435138012	Cocok
j2	14.1836843	13	0.007793732	0.361036425	Cocok
j3	6.218481601	15	0	0.975839395	Cocok
j5	23.30930477	16	0.017457306	0.105712921	Cocok
j6	17.49819464	15	0.010540633	0.289964342	Cocok
j7	17.94435385	15	0.011443228	0.265602352	Cocok
j8	23.84161017	16	0.018081806	0.093021433	Cocok
j9	19.96273116	15	0.014856412	0.173367034	Cocok
j10	20.71201812	15	0.015938533	0.14628551	Cocok
j11	5.987144638	13	0	0.946619569	Cocok
j12	23.06610873	17	0.015428717	0.147110837	Cocok
j13	15.22507608	16	0	0.508219481	Cocok
j14	34.0224457	15	0.029086183	0.003380738	Tidak cocok
j15	6.908371671	15	0	0.960142613	Cocok
j17	18.00757698	16	0.009149035	0.323453458	Cocok
j18	26.92351705	14	0.024815643	0.01969912	Tidak cocok
j19	19.89076606	15	0.014748301	0.176163206	Cocok
j20	14.38002377	15	0	0.49692431	Cocok
j21	28.93945368	16	0.023227216	0.024347588	Tidak cocok
j22	16.16332587	16	0.002609556	0.441621114	Cocok
j25	18.84939328	16	0.010899721	0.276550505	Cocok
j27	11.06638597	16	0	0.805369203	Cocok
j28	28.72406942	17	0.021449339	0.037156742	Tidak cocok
j29	30.33501376	16	0.024447714	0.016340666	Tidak cocok
j30	37.33356738	16	0.029824346	0.001881735	Tidak cocok

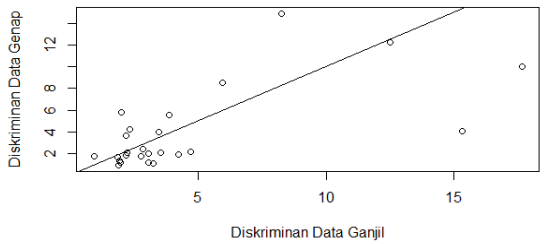
Pada Tabel 4 di atas, uji kecocokan model 4PL diperoleh sebanyak 19 butir yang cocok.



Gambar 1. Scree Plot Nilai Eigen untuk Hasil Ujian Akhir Semester Bahasa Inggris

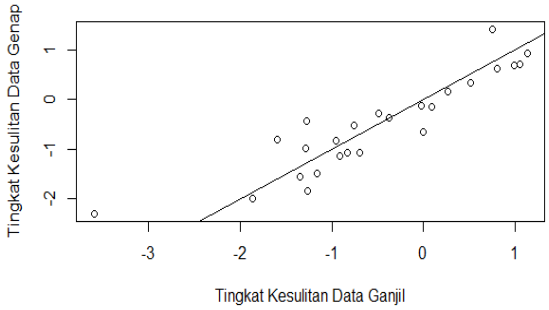
Asumsi unidimensi terpenuhi apabila pada *scree plot* terdapat satu komponen dominan dengan curaman tajam dan yang lain landai.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649



Gambar 2. Plot Invariansi Parameter *a*

Pada Gambar 2 terlihat bahwa titik-titik mengikuti garis lurus, walaupun ada beberapa titik yang melebar, akan tetapi dapat disimpulkan bahwa asumsi invariansi parameter *a* (diskriminan) terpenuhi.



Gambar 3. Plot Invariansi Parameter *b*

Terlihat pada gambar 3 bahwa titik-titik mengikuti garis lurus, dapat disimpulkan bahwa asumsi invariansi parameter *b* (tingkat kesulitan) terpenuhi. Pada pengujian invariansi parameter kemampuan (θ), data butir-butir hasil jawaban sebanyak 25, dibagi menjadi dua perangkat tes. Butir-butir hasil jawaban nomor ganjil menjadi perangkat tes pertama, dan butir-butir hasil jawaban nomor genap menjadi perangkat tes kedua. Setelah itu, dilakukan estimasi model 4PL pada kedua perangkat tes tersebut, sehingga diperoleh nilai kemampuan (*latent trait)* untuk perangkat tes bernomor ganjil dan untuk perangkat tes bernomor genap.

Tabel 5. Hasil Nilai Korelasi Hasil Ujian Akhir Semester Bahasa Inggris

	j1	j2	j3	j4	j5	j6
j1	1,000	0,23	0,16	-0,03	0,08	0,19
j2	0,23	1,000	0,3	-0,06	0,10	0,31
j3	0,16	0,3	1,000	-0,02	0,15	0,25
j4	-0,03	-0,06	-0,02	1,000	-0,04	0,03
j5	0,08	0,10	0,15	-0,04	1,000	0,11
j6	0,19	0,31	0,25	0,03	0,11	1,000

Berdasarkan hasil korelasi Tabel 5, menunjukkan bahwa nilai korelasi pada hasil ujian akhir semester bahasa Inggris kurang dari 0,50, sehingga dapat disimpulkan bahwa asumsi independensi lokal terpenuhi.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

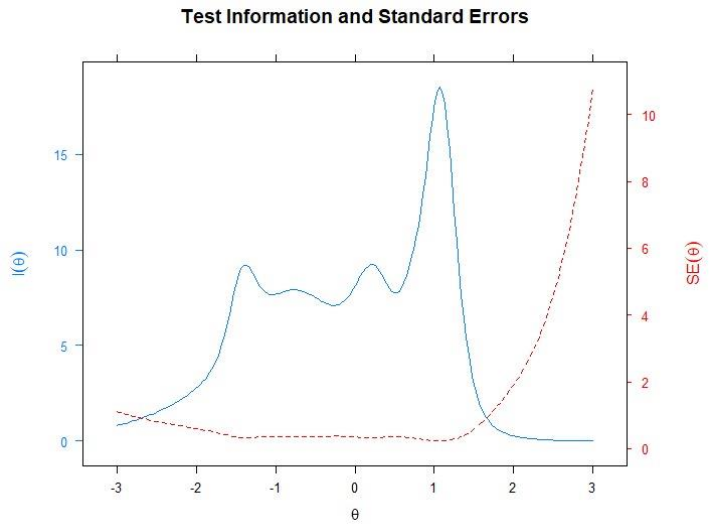
Berdasarkan model pada teori respon butir yang digunakan, dalam penelitian ini menggunakan model 4PL sebagai model yang paling cocok. Estimasi karakteristik butir soal pada perangkat tes menurut model 2PL adalah seperti pada tabel berikut.

Tabel 6. Karakteristik Butir Berdasarkan Model Terbaik

Butir	a	b	g	u
j1	1,551	-1,639	0,009	0,942
j2	2,982	-0,617	0,342	1,000
j3	2,171	-0,361	0,351	0,991
j5	3,515	0,961	0,338	0,998
j6	1,523	-0,897	0,001	0,952
j7	1,564	-1,242	0,035	0,998
j8	1,279	-1,226	0,002	0,943
j9	8,202	1,076	0,184	0,998
j10	2,501	-0,068	0,294	0,980
j11	2,472	-0,902	0,422	1,000
j12	5,977	0,803	0,115	0,955
j13	7,426	-1,473	0,557	0,979
j14	5,405	0,887	0,115	0,955
j15	2,666	-1,041	0,041	0,930
j17	2,711	-1,335	0,112	0,841
j18	1,803	-0,413	0,135	0,997
j19	1,570	-1,913	0,002	0,965
j20	2,024	-0,408	0,268	1,000
j21	3,134	0,373	0,214	0,903
j22	5,548	0,217	0,249	0,943
j25	8,563	0,701	0,161	0,276
j27	4,505	-0,591	0,153	0,347
j28	3,992	-0,012	0,233	0,685
j29	2,556	-0,793	0,337	0,847
j30	3,516	-0,948	0,453	0,910

Berdasarkan hasil pada Tabel 6, dapat dilihat bahwa nilai daya beda pada j1, j6, j7, j8, j18, j19 adalah baik. Sedangkan tingkat kesulitan butir pada semua butir baik. Sedangkan, pada indeks *pseudo guessing* pada j1, j6, j7, j15, j19 adalah baik. Dan pada indeks kecerobohan pada semua butir dikatakan baik kecuali pada butir j2, j11, dan j20.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649



Gambar 4. Kurva Nilai Fungsi Informasi dan SEM

Plot nilai fungsi informasi dan *standard error of measurement* (SEM) di atas menunjukkan bahwa garis lurus berupa nilai fungsi informasi dari kemampuan negative tak hingga mengalami kenaikan hingga mencapai nilai maksimum, kemudian menurun hingga positif tak hingga. Sedangkan, kesalahan pengukuran baku (SEM) yang bersimbol garis putus-putus adalah sebaliknya. Pada kurva antara nilai fungsi informasi dan SEM terdapat titik potong yang dihasilkan dari kedua garis tersebut yang berada diskala kemampuan (θ) pada interval -2,8 hingga +1,8. Pada interval tersebut, perangkat hasil ujian tulis bahasa Inggris memiliki nilai fungsi informasi yang lebih tinggi daripada SEM. Sebaliknya, apabila hasil ujian tulis bahasa Inggris diberikan pada peserta tes dengan kemampuan di luar rentang -2,8 hingga +1,8 akan memberikan nilai SEM yang lebih besar.

Tabel 7. Kategori Kemampuan Hasil Ujian Tulis

Interval Kemampuan	Kategori	Jumlah	Presentase (%)
$-4{,}00 \leq \theta < -2{,}00$	Sangat Rendah	81	5,4%
$-2{,}00 \leq \theta < -1{,}00$	Rendah	288	19,2%
$-1{,}00 \leq \theta < 1{,}01$	Rata-Rata	561	37,4%
$1{,}01 \leq \theta < 2{,}01$	Tinggi	551	36,7%
$2{,}01 \leq \theta \leq 4{,}00$	Sangat Tinggi	19	1,3%
Total		1500	100

Berdasarkan Tabel 7 menunjukkan bahwa kemampuan hasil ujian tulis akhir semester bahasa Inggris dapat diketahui 5,4% peserta tes tergolong ke dalam kategori sangat rendah dalam hal kemampuan hasil ujian tulis, 19,2% tergolong ke dalam kategori kemampuan hasil ujian tulis yang rendah, 37,4% tergolong ke dalam kategori kemampuan hasil ujian tulis yang rata-rata, 36,7% tergolong ke dalam kategori kemampuan hasil ujian tulis yang tinggi, dan 1,3% tergolong ke dalam kategori kemampuan hasil ujian tulis yang sangat tinggi. Dengan demikian, dapat disimpulkan bahwa persebaran kemampuan hasil ujian tulis akhir semester Bahasa Inggris secara keseluruhan masuk ke dalam kategori rata-rata.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Pembahasan

Penelitian ini bertujuan untuk membandingkan kecocokan model dengan pendekatan teori respons butir penskoran dikotomi pada data hasil ujian tulis yang menempuh bahasa Inggris niaga pada ujian akhir semester perguruan tinggi se-Indonesia pada tahun 2022 serta menentukan model yang terbaik yang akan digunakan pada penelitian ini. Model penskoran dikotomi yang digunakan antara lain *Rasch*, 1PL, 2PL, 3PL, dan 4PL. Pada pengujian kecocokan model dilakukan analisis secara statistik dengan menggunakan metode *Yen's Q1,* serta memperhatikan nilai AIC dan BIC. Selain itu, dilakukan analisis karakteristik butir hasil jawaban dengan menggunakan model terbaik yang diperoleh, serta mampu mendeskripsikan tingkat kemampuan bahasa Inggris di Indonesia. Teori respon butir memiliki beberapa asumsi yang harus dipenuhi diantaranya asumsi unidimensi, asumsi invariansi parameter, dan asumsi independensi lokal. 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 menandakan bahwa asumsi invariansi parameter pada penelitian ini dapat terpenuhi ketika asumsi unidimensi dapat terpenuhi.

Hasil dari uji kecocokan model menggunakan metode *Yen's Q1* dengan model 4PL sebagaimana pada Tabel 14 terdapat 19 butir yang cocok dan 6 butir yang tidak cocok dimana terdapat pada butir 14, butir 18, butir 21, butir 28, butir 29, dan butir 30. Hal ini sebagaimana pada Tabel 12, juga dikatakan cocok sebanyak 19 butir dan 6 butir yang tidak cocok dimana terdapat pada butir 14, butir 18, butir 21, butir 28, butir 29, dan butir 30. Sedangkan pada Tabel 10 ada sebanyak 11 butir yang cocok yang terdapat pada butir 1, butir 2, butir 3, butir 6, butir 7, butir 8, butir 11, butir 19, butir 20, butir 27, dan butir 27. Dan pada Tabel 8 dan Tabel 6 terdapat 6 butir yang cocok, dimana pada Tabel 8 terdapat pada butir 1, butir 3, butir 5, butir 7, butir 12, dan butir 17. Sedangkan pada Tabel 6 terdapat pada butir 1, butir 3, butir 6, butir 8, butir 13, dan butir 19. Dari kelima model tersebut, kemudian dibandingkan nilai AIC dan BIC. Sebagaimana terdapat pada Tabel 16 menunjukkan bahwa nilai AIC dan BIC terkecil terdapat pada model 4PL, dimana nilai AIC sebesar 37967,02 dan nilai BIC sebesar 38498,34. Sehingga, model terbaik yang digunakan pada penelitian ini adalah model 4PL. Penelitian yang selaras terkait metode 4PL juga pernah dilakukan oleh Jimoh Kasali, Adediwura Alaba Adeyemi (2022) dimana hasil penelitian menyimpulkan bahwa model logistik empat parameter cocok dengan soal tes ujian nasional matematika. Serta penelitian sebelumnya yang selaras dengan hasil ini pernah dilakukan oleh Lucy Barnard-Brak dan Zhanxia Yang (2023) dimana nilai AIC, BIC yang lebih rendah mengindikasikan kecocokan model yang lebih baik satu sama lainnya. Grafik pada plot ICC yang dihasilkan oleh model 4PL juga termasuk dalam kategori baik dan dapat diterima karena mengikuti ogif normal meskipun ada beberapa butir diantaranya tidak mengikuti bentuk orgif normal. Penelitian yang dilakukan oleh Timbul (2023) menunjukkan bahwa semakin sulit suatu soal maka posisi titik belok kurva sifat soal tersebut semakin kekanan atau semakin besar nilai theta dituntut untuk mampu memberikan respon yang benar terhadap suatu item.

Hasil ini selaras dengan penelitian yang telah dilakukan oleh Aslam Fatkhudina, et al. (2014) pada penelitiannya diperoleh hasil bahwa aplikasi CAT yang dipadukan dengan model IRT 4PL dapat mengukur kemampuan tes lebih pendek atau lebih cepat dan juga peluang menjawab dengan benar soal tes yang dikerjakan cenderung lebih baik daripada model IRT

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

3PL. Pada penelitian yang lain yang dilakukan oleh Omur Kaya Kalkan (2020) yang dimana model IRT 4PL dan DINA dapat dipertimbangkan untuk menganalisis kumpulan data yang terkontaminasi dengan efek *pseudo guessing* dan *slipping*. Hal ini menunjukkan bahwa model terbaik yang dihasilkan bergantung pada data yang digunakan. Dimana setiap data yang dimiliki terdapat karakteristiknya masing-masing, seperti data yang kompleks serta asumsi dasar yang berbeda. Selain itu ukuran sampel dan karakteristik sampel yang berbeda juga menjadi suatu hal yang berpengaruh terhadap pemilihan model terbaik. Contohnya, pada penelitian ini ukuran sampel yang digunakan sebanyak 1500 responden dengan karakteristik mahasiswa semester genap pada sebuah perguruan tinggi di Indonesia, sedangkan pada penelitian yang dilakukan oleh Aslam Fatkhudina, et al. (2014) tersebut menggunakan sebanyak 172 responden dengan karakteristik siswa yang tersebar di 6 kelas pada UAS Bahasa Inggris. Selain itu, model 4PL juga menggunakan empat parameter, yaitu nilai daya beda, tingkat kesulitan, tebakan semu, dan kecerobohan sehingga dalam mengukur suatu kemampuan peserta tes dapat lebih akurat. Kemudian pada penelitian yang dilakukan Alexander Robitzsch (2022) dataset yang disimulasikan terdiri dari 30 item dimana ukuran sampel dari set data respons butir sebagai N= 1000, 2000, 5000, dan 10.000 untuk mencerminkan situasi yang berbeda namun umum terjadi pada aplikasi data tes pendidikan. Dan pada penelitian yang dilakukan oleh Faye Antoniou, Ghadah Alkhadim, Angeliki Mouzaki, dan Panagiotis Simos (2022) partisipan yang digunakan sejumlah 1127 anak berusia 5 hingga 11 tahun.

Pada analisis karakteristik butir yang dilakukan dengan model 4PL sebagai model terbaik yang dimana memberikan hasil sebagaimana terdapat pada Tabel 19 yang mengklarifikasikan butir soal berdasarkan nilai daya beda menjadi butir jawaban yang cukup pada j1, j6, j7, j8, j18, j19 adalah baik. Sedangkan tingkat kesulitan butir pada semua butir baik. Sedangkan, pada indeks *pseudo guessing* pada j1, j6, j7, j15, j19 adalah baik. Dan pada indeks kecerobohan pada semua butir dikatakan baik kecuali pada butir j2, j11, dan j20. Hal ini menunjukkan bahwa jika suatu perangkat tes memiliki nilai daya beda dan tingkat kesulitan yang baik maka perangkat tes tersebut dapat memberikan informasi yang baik mengenai kemampuan peserta tes secara keseluruhan. Hasil ini selaras dengan penelitian yang dilakukan oleh Duden Saepuzaman, Edi Istiyono, Haryanto, Heri Retnawati, Yustiandi (2021) dimana hasil penelitian menunjukkan bahwa parameter butir soal Fisika materi Usaha dan Energi dengan penskoran dikotomus menunjukkan keseluruhan butir memiliki kriteria butir baik dan dengan penskoran politomus menunjukkan hamper seluruh butir kategori baik.

Berdasarkan nilai fungsi informasi dan SEM diperoleh hasil seperti pada Gambar 12 dan Tabel 20 bahwa diperoleh nilai fungsi informasi tertinggi sebesar 18,51 pada kemampuan (θ) sebesar 1,07 dengan tingkat kesalahan pengukuran baku (SEM) sebesar 1. Selain itu, nilai fungsi informasi akan lebih tinggi daripada SEM ketika kemampuan θ berkisar -2,8 hingga +1,8. Hal ini dapat disimpulkan bahwa hasil ujian tulis mampuan memberikan informasi mengenai kemampuan peserta tes dalam kategori sangat rendah hingga tinggi. Penelitian yang selaras dengan hasil fungsi nilai informasi tersebut pernah dilakukan oleh Reza Oktiana Akbar (2021) dimana hasil analisis menunjukkan rata-rata tingkat kesulitan versi dikotomi sebesar 0,166 dengan simpangan baku sebesar 1,137, sedangkan rata-rata tingkat kesulitan versi politomi sebesar 0,033 dengan simpangan baku sebesar 0,940. Nilai fungsi informasi versi penskoran dikotomi lebih tinggi dibandingkan dengan penskoran versi politomi.

Kemudian, berdasarkan estimasi parameter kemampuan yang dilakukan menggunakan metode estimasi Bayes dengan EAP mendapatkan hasil sebagaimana pada Gambar 13 yang menunjukkan bahwa sebaran kemampuan peserta dalam mengerjakan soal ujian tulis dengan 4PL terbanyak mendekati mean, yakni berada di sekitar nol. Selain itu, dijelaskan lebih rinci seperta pada Tabel 22 bahwa dengan 1500 peserta tes diperoleh rata-rata kemampuan hasil ujian

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

tulis sebesar 0. Hal ini dapat dikatakan bahwa secara keseluruhan peserta tes memiliki kemampuan hasil ujian tulis pada kategori rata-rata, dimana sebesar 37,4% dari total jumlah peserta tes memiliki kemampuan pada rentang -1,00 hingga 1,01.

SIMPULAN

Berdasarkan hasil perbandingan kecocokan model dan analisis karakteristik butir pada hasil ujian tulis menggunakan pendekatan teori respons butir penskoran dikotomi, dapat disimpulkan sebagai berikut:

1. Kecocokan model *Rasch* pada data soal ujian tulis bahasa Inggris yang diskor dikotomi diperoleh sebanyak 6 butir yang cocok. Nilai AIC dan BIC pada model *Rasch* yaitu 38891,69 dan 39029,83.
2. kecocokan model 1PL pada data soal ujian tulis bahasa Inggris yang diskor dikotomi diperoleh sebanyak 6 butir yang cocok. Nilai AIC dan BIC pada model 1PL yaitu 38891,69 dan 39029,84.
3. kecocokan model 2PL pada data soal ujian tulis bahasa Inggris yang diskor dikotomi diperoleh sebanyak 11 butir yang cocok. Nilai AIC dan BIC pada model 2PL yaitu 38378,28 dan 38643,94.
4. kecocokan model 3PL pada data soal ujian tulis bahasa Inggris yang diskor dikotomi diperoleh sebanyak 19 butir yang cocok. Nilai AIC dan BIC pada model 3PL yaitu 38082,75 dan 38481,25.
5. kecocokan model 4PL pada data soal ujian tulis bahasa Inggris yang diskor dikotomi diperoleh sebanyak 19 butir yang cocok. Nilai AIC dan BIC pada model 4PL yaitu 37967,02 dan 38498,34. Pada model 4PL nilai AIC dan BIC menjadi nilai yang paling rendah diantara model yang lainnya.
6. Berdasarkan hasil yang telah diuraikan pada poin 1 sampai 5, diperoleh model yang terbaik untuk digunakan adalah model 4PL.
    a. Analisis karakteristik butir hasil ujian tulis dengan model 4PL didapatkan hasil bahwa nilai daya beda, tingkat kesulitan, tebakan semu, dan kecerobohan tergolong baik. Tingkat kesulitan pada suatu butir akan semakin bertambah besar seiring dengan bertambahnya kategori pada masing-masing butir.
    b. Hasil analisis menggunakan model 4PL menunjukkan hasil bahwa nilai fungsi informasi yang maksimum sebesar 18,51 pada kemampuan (θ) sebesar 1,07 dengan tingkat kesalahan pengukuran baku (SEM) sebesar 1. Selain itu, nilai fungsi informasi akan lebih tinggi daripada SEM ketika kemampuan θ berkisar -2,8 hingga +1,8.
    c. Tingkat kemampuan hasil ujian tulis yang menempuh bahasa Inggris niaga pada ujian akhir semester perguruan tinggi se-Indonesia secara keseluruhan berada pada kategori rata-rata, dimana peserta tes diketahui 5,4% peserta tes tergolong ke dalam kategori sangat rendah dalam hal kemampuan hasil ujian tulis, 19,2% tergolong ke dalam kategori kemampuan hasil ujian tulis yang rendah, 37,4% tergolong ke dalam kategori kemampuan hasil ujian tulis yang rata-rata, 36,7% tergolong ke dalam kategori kemampuan hasil ujian tulis yang tinggi, dan 1,3% tergolong kedalam kategori kemampuan hasil ujian tulis yang sangat tinggi.

UCAPAN TERIMA KASIH

Puji syukur atas kehadirat Allah yang telah melimpahkan rahmatNya, sehingga penulis dapat menyelesaikan Tugas Akhir Skripsi dengan lancar. Dengan hati yang tulus, penulis mempersembahkan karya tulis ini kepada:

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

1. Keluarga tercinta yaitu Kakek Narja, Nenek Suwarni, Tante Lia, Om Yono, Adik Rara, Adik Rani, dan Adik Rania yang selalu memberikan doa, semangat, dan motivasi kepada saya hingga saat ini dan seterusnya. Penulis mendoakan semoga segala kebaikan yang diberikan kepada penulis mendapat balasan dari Allah SWT.
2. Papa Haryanto, Mama Icha, Mama Willa Ferbrianti, dan Papa Ony. Selaku orang tua yang jauh di Pontianak sana yang selalu mendoakan, memberikan semangat, dan motivasi untuk menyelesaiakan Tugas Akhir Skripsi, semoga diberikan kesehatan dan rezeki yang berlimpah.
3. Tanti, Jundi, Lala, Silvi yang selalu meluangkan waktunya untuk main, diskusi, dan menghibur, dan mengingatkan penulis setiap hari.
4. Ratna Tri Utami yang selalu meminjamkan laptopnya dikala laptop penulis bermasalah di aplikasi pengolah data.
5. Teman-teman Karang Taruna Nyawiji Padukuhan Blumbang
6. Temen-teman Orsen Himpunan Mahasiswa Matematika dan Badan Eksekutif Mahasiswa
7. Teman kelas Statistika angkatan 2019
8. Semua pihak yang telah membantu secara langsung maupun tidak langsung sehingga Tugas Akhir Skripsi dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

Akbar, R. O. (2021). *Comparative Analysis of Question Item Parameters and Students' Ability between Dichotomy and Polytomic Score Versions; Research on Mathematics National Exam Test Participants*. http://i-jeh.com/index.php/ijeh/index

Antoniou Faye, Alkhadim G., Mouzaki A., Simos P. (2022). *A psychometric analysis of raven's colored progressive matrices: Evaluating guessing and carelessness using the 4pl item respons theory model.* Journal of intelligence. https://doi.org/10.3390/jintelligence100100006

Brak, Lucy. B. (2023). *A 4pl item response theory examination of perceived stigma in the screening of eating disorders*. https://doi.org/10.1007/s40519-023-01604-w

Anisa (2013). Pengaruh penggunaan web dengan software berbasis open source terhadap peningkatan berpikir kreatif siswa (Skripsi UPI, Bandung).

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software, 48(6), 1-29.

Fatkhudin, A., Surarso, B., & Subagio, A. (2014). Item response theory model empat parameter logistik pada computerized adaptive test. *Jurnal Sistem Informasi Bisnis.*

Hambelton, R.K. (1991). *Fundamentals of item response theory*. California: Sage Publications, Inc.

Kalkan, Ö. K., & Çuhadar, İ. (2020). An evaluation of 4pl irt and dina models for estimating pseudo-guessing and slipping parameters. *Journal of Measurement and Evaluation in Education and Psychology*, *11*(2), 131–146. https://doi.org/10.21031/epod.660273.

Kasali, J., & Adeyemi, A. A. (2022). Model-data fit using akaike information criterion (aic), bayesian information criterion (bic), and the sample-size-adjusted bic. *Square : Journal of Mathematics and Mathematics Education*, *4*(1), 43–51. https://doi.org/10.21580/square.2022.4.1.11297.

Le, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. Journal of Statistical Software, 25(1).

Paek, I., & Cole, K. (2020). Using R for item response theory model applications. Routledge.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Pardede, T., Sabtoso, A., Diki, Retnawati, H., Rafi, I., Apino, E., Rosyada, M. N. (2023). Gaining a deeper understanding of the meaning of the carelessness parameter in the 4PL IRT model and strategies for estimating it.

Robitzsch, A. (2022). *Four parameter guessing model and related item response models.* https://doi.org/10.3390/mca27060095

Retnawati, H. (2014). Teori respons butir dan penerapannya: Untuk penelitian, praktisi pengukuran dan pengujian, mahasiswa pascasarjana. Yogyakarta: Nuha Medika.

Saepuzaman, D., Istiyono, E., Haryanto, Retnawati, H., Yustiandi. (2021). Analisis Karakteristik Butir Soal Fisika Dengan Pendekatan IRT Penskoran Dikotomus dan Politomus. https://doi.org/10.37729/radiasi.v14i2.1200

Sudaryono (2011). Implementasi teori respon butir (item response theory) pada penilaian hasil belajar akhir sekolah. *Jurnal Pendidikan dan Kebudayaan*, 17(6).

Noventa, S., Ye, S., Kelava. A. (2024). *On the identifiability of 3- and 4-parameter item response theory models from the perspective of knowledge space theory*. https://doi.org/10.1007/s11336-024-09950-z

Loken, E., & Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 509–525. https://doi.org/10.1348/000711009X474502.

Copyright © 2024, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649