

Uji Kecocokan Model Parameter Logistik Soal Diagnosa Kemampuan Matematika Dasar

Wahyu Hartono^{1*}, Samsul Hadi², Raden Rosnawati³, Heri Retnawati⁴

^{1*}Pendidikan Matematika, Universitas Swadaya Gunung Jati, Cirebon, Indonesia;

^{1*}Program Doctoral Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; ^{1*}wahyuhartono.2019@student.uny.ac.id

²Program studi Pendidikan Teknik Elektro, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; ²samsul.hd@gmail.com

^{3,4}Program studi Pendidikan Matematika, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia; ³rosnawati_uny@yahoo.com; ⁴heri_retnawati@uny.ac.id

Info Artikel: Dikirim: 06 Oktober 2021; Direvisi: 13 Januari 2022; Diterima: 27 Januari 2022

Cara sitasi: Hartono, W., Hadi, S., Rosnawati, R., & Retnawati, H. (2022). Uji Kecocokan Model Parameter Logistik Soal Diagnosa Kemampuan Matematika Dasar. *JNPM (Jurnal Nasional Pendidikan Matematika)* 6(1), 125-144.

Abstrak. Penilaian diagnostik kognitif berfungsi untuk mengukur struktur pengetahuan teknis siswa dan kemampuan pemrosesan untuk memberikan informasi tentang kekuatan dan kelemahan kognitifnya. Penelitian ini bertujuan untuk menganalisis kesesuaian model teori respon butir soal tes diagnostik kemampuan matematika dasar. Jenis penelitian ini adalah penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Subyek penelitian sebanyak 200 responden yang merupakan mahasiswa mahasiswa tingkat 1,2,3 dan 4 dari 7 program sarjana Universitas Swadaya Gunung Jati. Instrumen yang digunakan berupa tes materi eksponensial berbentuk multiple-choice sebanyak 12 butir. Data yang diperoleh berupa respon siswa pada instrumen tes yang diberikan. Teknik analisis datanya adalah analisis karakteristik butir berdasarkan teori respons butir (IRT). Analisis uji prasyarat model IRT dengan analisis faktor. Uji asumsi independensi lokal berdasarkan penghitungan korelasi antar item. Sedangkan uji invariansi parameter menggunakan *scatter plot*. Berdasarkan hasil penelitian ini diperoleh nilai korelasi poin biserial butir nomor empat adalah negatif, sehingga butir tersebut tidak digunakan dalam analisis parameter dan kecocokan model. Model teori respons butir yang cocok dengan instrumen tersebut adalah model 3-PL. Tingkat kesukaran item berada dalam rentang -0.918 sampai 0.901 . Daya pembeda berada dalam rentang $1,134$ sampai $21,161$ serta tebakan semu ada di antara 0 sampai 0.387 . Jika ditinjau dari tingkat kesulitan butir terbukti bahwa keseluruhan butir memiliki tingkat kesulitan yang baik, sehingga dapat disimpulkan bahwa seluruh butir memiliki karakteristik butir yang baik. Penelitian ini memberikan gambaran pengujian empiris dalam menentukan kualitas butir instrumen berdasarkan analisis yang tepat sehingga layak untuk digunakan dalam mengukur hasil belajar.

Kata Kunci: Tes Diagnosa, Teori Respons Butir, Model Logistik Parameter.

Abstract. Cognitive diagnostic assessment serves to measure the structure of students' technical knowledge and processing abilities to provide information about their cognitive strengths and weaknesses. This study aims to analyze the suitability of the response theory model for basic mathematical ability diagnostic test items. This type of research is descriptive exploratory research with a quantitative approach. The research subjects were 200 respondents who were students of levels 1,2,3,

and 4 from 7 undergraduate programs at the Swadaya Gunung Jati University. The instrument used is an exponential material test in multiple-choice as many as 12 items. The data was obtained in student responses to the given test instrument. The data analysis technique is item characteristic analysis based on item response theory (IRT). Prerequisite test analysis of item response theory model with factor analysis. Test the assumption of local independence based on calculating the correlation between items. While the parameter invariance test uses a scatter plot. Based on the results of this study, the biserial point correlation value of item number four is negative, so that item is not used in the parameter analysis and model fit. The item response theory model that fits the instrument is the 3-PL model. The item difficulty level is in the range of -0.918 to 0.901. Distinguishing power is in the range of 1.134 to 21,161, and false guesses are 0 to 0.387. If viewed from the level of difficulty of the items, it is proven that all items have a good level of difficulty, so it can be concluded that all items have good item characteristics. This study provides an overview of how empirical testing determines the quality of instrument items based on appropriate analysis so that it is feasible to use in measuring learning outcomes.

Keywords: Diagnostic Tests, Item Response Theory, Parameter Logistics Model.

Pendahuluan

Matematika adalah pelajaran yang terurut, artinya siswa dapat menguasai level matematika yang lebih tinggi jika telah menguasai level yang lebih rendah. Penguasaan kemampuan prasyarat adalah hal yang penting untuk dipenuhi. Menurut Levine, Lindsay, & Reed (1992), matematika secara intrinsik bersifat kumulatif. Hirarki pengetahuan dan keterampilan harus dibangun dari waktu ke waktu. Ini berarti bahwa siswa yang penguasaan awal subjek matematika tidak lengkap cenderung sulit untuk mengikuti materi berikutnya. Memori bersifat kumulatif, oleh karena itu, merupakan subkomponen penting dari kemampuan matematika. Seharusnya siswa benar-benar menguasai kemampuan matematika ditingkat Sekolah Menengah Atas agar dapat mengikuti pelajaran matematika di tingkat Universitas. Menurut teori belajar Robert Gagne (Tall & Razali, 1993), empat fase belajar adalah fase menangkap (apprehending), fase akuisisi (acquisition), fase menyimpan (storage), dan fase mengambil (retrieval). Siswa dengan kemampuan tinggi dapat melewati keempat fase tersebut. Tetapi bagi siswa dengan kemampuan rendah akan kesulitan melewati fase keempat (retrieval).

Salah satu upaya untuk membantu keberhasilan belajar matematika siswa adalah dengan mendiagnosa kesulitan belajarnya. Hal ini sangat rasional karena hasil diagnostic ini dapat dijadikan pertimbangan untuk menarik kesimpulan atau keputusan dalam upaya untuk meningkatkan kemungkinan memberikan instruksi yang paling efektif dan tepat waktu kepada siswa dengan memotong langsung ke sumber masalahnya. Sehingga dengan informasi tersebut, guru dapat mengubah kesalahpahaman siswa dan mengganti strategi yang salah (Leighton & Gierl, 2007).

Salah satu unsur yang harus diperhatikan dalam pelaksanaan tes diagnosa adalah mengusahakan dan memastikan hasil tes tersebut menggambarkan kemampuan siswa secara akurat. Suatu penilaian disebut akurat jika hasil penilaian mengandung kesalahan atau *error* sekecil mungkin. Untuk mendapatkan hasil yang akurat menggambarkan kemampuan peserta didik, kualitas instrumen tes dalam hal ini soal-soal tes diagnosa matematika dasar harus valid, reliabel dan memiliki parameter item yang baik.

Pengujian instrumen tes diagnostik banyak dilakukan dalam studi-studi terdahulu. melakukan studi pengujian instrumen tes diagnostik untuk analisis pemahaman siswa pada materi kelarutan dan hasil kali kelarutan. Pengujian kualitas instrumen didasarkan pada hasil validasi ahli mengenai kelayakan instrumen mendapat rerata skor 36,8 dari jumlah skor total 44 dengan kategori sangat layak dan reliabilitas sebesar 0,72. mengembangkan tes untuk mendiagnosis letak kesulitan belajar siswa khususnya pada materi bangun ruang sisi lengkung. Pengujian kualitas instrument ditinjau dari validitas ahli dan rekiabilitas instrumen secara empirik. mengembangkan tes diagnostic untuk mendapatkan deskripsi kesulitan siswa dalam belajar matematika pada pokok bahasan bangun datar, ditinjau dari segi kemampuan koneksi matematika siswa. Instrumen yang digunakan diuji kualitasnya berdasarkan validitas konten berdasarkan masukan ahli. Penelitian-penelitian terdahulu ada kecenderungan penentuan kualitas instrument hanya didasarkan pada validtas konten dan reliabilitas instrument tes secara keseluruhan. Selain aspek validitas dan reliabilitas, kualitas instrumen juga ditinjau dari karakteristik butir yang digunakan. Karakteristik butir yang dimaksud disini adalah parameter-parameter yang dimiliki butir seperti daya beda, tingkat kesulitan dan tebakan semu.

Untuk tujuan ini, ada dua jenis pendekatan yang dapat digunakan untuk memperkirakan parameter item, yaitu teori tes klasik dan teori respons butir. Tes klasik Teori dipandang memiliki kelemahan. Kelemahan yang paling menonjo dari teori tes klasik adalah bahwa karakteristik peserta ujian dan karakteristik tes tidak dapat dipisahkan, masing-masing dapat hanya ditafsirkan dalam konteks lain (Hambleton, Swaminathan, & Rogers ,1991). Artinya, kemampuan peserta ujian hanya ditentukan oleh tes. Ketika Tes sulit, peserta ujian akan tampak memiliki kemampuan yang rendah, dan ketika ujian mudah, peserta ujian akan tampak memiliki kemampuan yang lebih tinggi. Dengan kata lain, parameter item sangat tergantung pada subjek/ peserta tes begitupun sebaliknya. Karakteristik item akan berubah ketika peserta ujian berubah, dan karakteristik peserta ujian akan berubah ketika

itemnya berubah. Dalam hal ini, teori tes klasik tidak dapat digunakan sebagai standar karena hasil penilaian sangat tergantung pada subyek peserta tes.

Teori respons butir merupakan solusi untuk mengatasi kelemahan yang ada pada teori tes klasik karena teori respons butir memiliki konsep melepaskan keterkaitan antara item dan sampel atau subyek peserta tes. Karakteristik/kemampuan peserta ujian tetap meskipun mereka bekerja pada item dengan karakteristik yang berbeda, dan sebaliknya, karakteristik item tetap meskipun mereka dikerjakan oleh peserta ujian dengan kemampuan yang berbeda. Selain itu juga, teori respon butir mendasarkan pada item/butir bukan lagi pada perangkat tes. Fungsi dari teori respons butir bisa diterapkan ketika model yang digunakan memiliki kecocokan dengan data uji (Hambleton, Swaminathan, & Rogers, [1991](#)). Stone & Zhang ([2003](#)) menyatakan bahwa penggunaan estimasi parameter butir bisa terganggu ketika model yang digunakan tidak cocok dengan data. Beberapa model logistik dalam teori respons butir, yaitu model satu parameter logistik 1-PL, model dua parameter logistik 2-PL, dan model tiga parameter logistik 3-PL (Hambleton, Swaminathan, & Rogers, [1991](#)). Setiap model memiliki perbedaan jumlah parameter butir. Parameter butir ini berfungsi sebagai pembentuk fungsi respons butir. Penggunaan teori respon butir model 1-PL, 2 PL dan 3 PL perlu memilih, data yang dianalisis apakah sesuai dengan salah satu dari ketiga model tersebut. Untuk menentukan kesesuaian model parameter logistik yang dipakai, setidaknya ada dua metode yang bisa dilakukan, yaitu metode statistik dan metode grafik (Retnawati, [2014](#)). Berdasarkan latar belakang, studi ini bertujuan untuk mendeskripsikan pengujian kualitas instrumen tes diagnostik kemampuan matematika dasar materi eksponensial dengan pendekatan teori respon butir.

Metode

Jenis penelitian ini adalah penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Partisipan penelitian sebanyak 200 responden yang merupakan mahasiswa tingkat 1, 2, 3 dan 4 dari 7 program sarjana Universitas Swadaya Gunung Jati. Responden yang dipilih adalah mahasiswa yang pernah mendapatkan mata kuliah matematika dasar.

Instrumen yang digunakan berupa tes materi eksponensial berbentuk *multiple-choice* sebanyak 12 butir. Pengembangan instrumen mengacu pada capaian pembelajaran mata kuliah yang dikembangkan dari capaian pembelajaran prodi.

Data yang digunakan adalah respon siswa pada instrument tes yang diberikan. Pelaksanaan pengumpulan data dilakukan secara ketat sehingga diharapkan respon siswa mencerminkan hasil belajar siswa yang sebenarnya.

Teknik analisis data yang digunakan adalah analisis karakteristik butir soal berdasarkan teori respons butir (IRT) dengan perbedaan model parameter logistik (PL), yakni model 1-PL, 2-PL, dan model 3-PL. Karakteristik butir soal untuk ketiga model parameter logistik tersebut berdasarkan nilai khi-kuadrat, dan *plot* kurva karakteristik (ICC) menggunakan *software R*. Analisis uji prasyarat model teori respon butir dengan analisis faktor eksplorasi (EFA) menggunakan SPSS untuk menentukan dimensi yang diukur dalam instrumen. Kemudian, uji asumsi independensi lokal berdasarkan penghitungan korelasi antar item serta invariansi parameter, yakni penghitungan parameter butir dan kemampuan menggunakan *software R* dan *plot* menggunakan *Excell*.

Hasil dan Pembahasan

Sebelum melakukan uji prasyarat unidimensionalitas data, independensi lokal, dan invariansi pengukuran, terlebih dahulu dilakukan analisis butir soal secara klasik dengan memperhatikan korelasi poin biserial untuk setiap butir yang digunakan dalam penelitian ini. Hasil penghitungan menggunakan *software R* diperoleh nilai korelasi poin biserial untuk setiap 12 butir soal. Nilai-nilai tersebut tersaji pada Tabel 1 berikut.

Tabel 1. Nilai Koefisien Poin Biserial

Item	Korelasi Poin Biserial
Item1	0.4057
Item2	0.7436
Item3	0.4957
Item4	-0.1963
Item5	0.6987
Item6	0.6262
Item7	0.6909
Item8	0.4620
Item9	0.2710
Item10	0.4434
Item11	0.6544
Item12	0.5670

Tabel 1 menunjukkan bahawa terdapat nilai korelasi poin biserial yang negatif untuk 1 butir soal, yakni butir soal nomor empat sehingga butir tersebut dieliminasi dan tidak digunakan untuk analisis kecocokan model teori respon

butir. Selanjutnya dilakukan uji prasyarat dalam menentukan analisis model respon butir yang digunakan untuk analisis kecocokan model.

Sebelum melakukan uji kecocokan model, langkah pertama adalah membuktikan asumsi teori respons butir, yaitu unidimensional, independensi lokal dan invarian parameter. Unidimensi berarti setiap item hanya mengukur satu kemampuan (Retnawati, 2014). Asumsi unidimensi dalam penelitian ini dibuktikan melalui analisis faktor menggunakan SPSS. Analisis faktor dilakukan dengan terlebih dahulu melakukan analisis uji kelayakan, yaitu KMO-MSA tes dan uji Barlett. Tes KMO-MSA bertujuan untuk melihat kecukupan sampel, sedangkan Uji Barlett berfungsi untuk membuktikan homogenitas data. Analisis faktor dapat dilanjutkan jika nilai Kaiser Meyer Olkin (KMO)-MSA $> 0,5$ dan uji signifikan Barlett $< 0,05$ (Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L., 2009). Berdasarkan data respon dalam studi ini dan dengan menghapus butir nomor empat dimana nilai korelasi poin biserialnya adalah negatif, diperoleh nilai KMO-SMA dan Barlett seperti disajikan pada Tabel 2.

Tabel 2. Uji KMO dan Bartlett's

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		
		0,884
Bartlett's Test of Sphericity		
Approx. Chi-Square		821,183
df		55
Sig.		,000

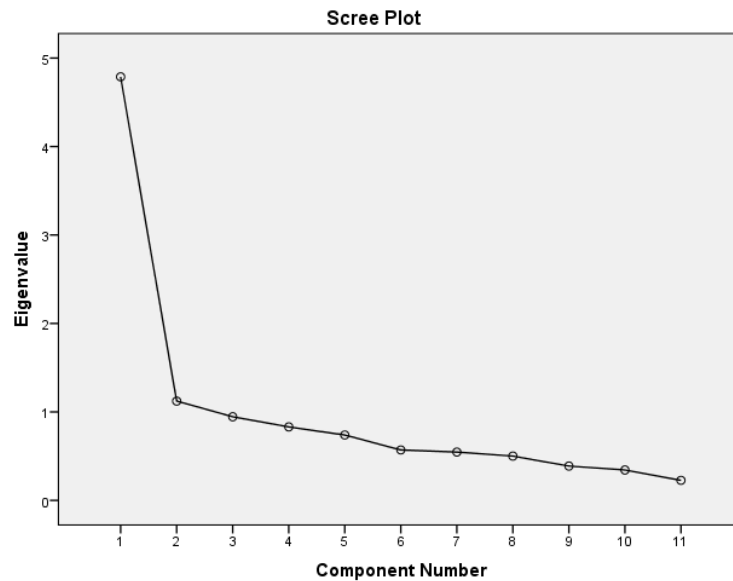
Tabel 2 menunjukkan nilai KMO-MSA adalah 0,884 dan signifikan uji Bartlett adalah 0,000. Ini berarti sampel yang digunakan telah memenuhi persyaratan kecukupan sampel dan datanya adalah data yang homogen, sehingga analisis faktor dapat dilakukan. Hasil pengolahan data untuk analisis faktor melalui SPSS dapat dilihat di bagian nilai Eigen pada Tabel 3.

Tabel 3. Nilai Eigen

Component	Initial Eigenvalues		
	Total	% of Varians	Cumulative %
1	4,788	43,528	43,528
2	1,123	10,205	53,733

Tabel 3 menunjukkan nilai Eigen yang nilainya lebih dari satu menandakan satu faktor, berdasarkan nilai tersebut, tes diagnosa kemampuan matematika dasar memiliki dua faktor. Dua faktor ini dapat menjelaskan 53,733% varians. Nilai Eigen tersebut kemudian dapat disajikan dalam *scree plot* pada Gambar 1. *Scree plot* dari analisis faktor menunjukkan penurunan yang sangat tajam antara faktor 1 dan faktor 2, nilai Eigen kemudian mulai miring pada faktor ke-3 sehingga *scree plot* hampir membentuk sudut siku kanan. Ini

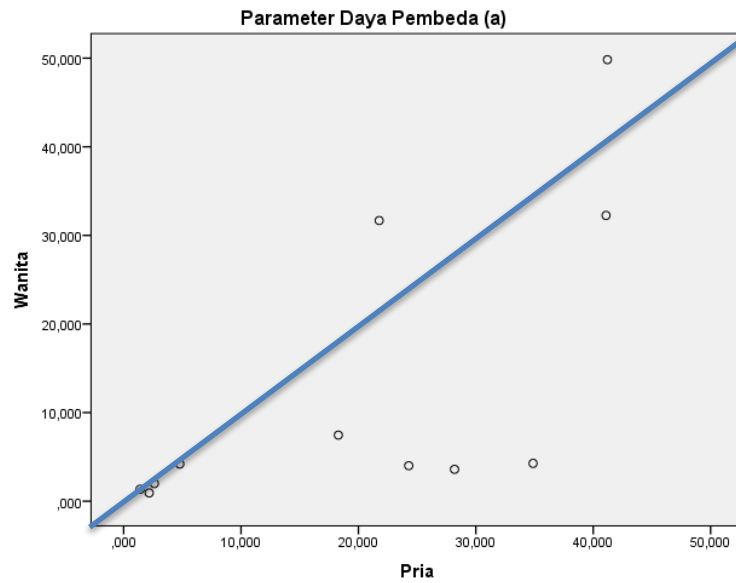
menunjukkan bahwa hanya ada 1 faktor dominan dalam perangkat tes sehingga asumsi unidimensional terpenuhi.



Gambar 1. Analisis Faktor *Scree plot*

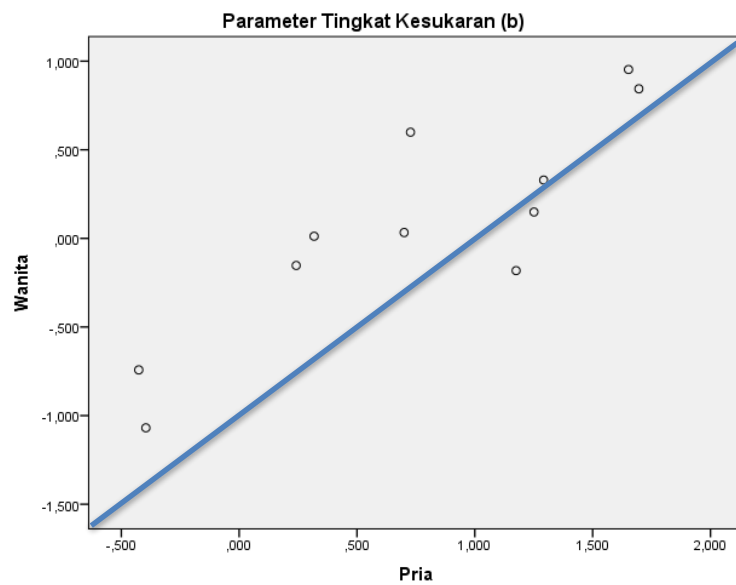
Faktor dominan tersebut adalah kemampuan penguasaan konsep materi eksponen. Menurut De Mars (2010), independensi lokal dapat terdeteksi pula dengan membuktikan asumsi unidimensional. Ini dapat diartikan bahwa jika asumsi unidimensional terpenuhi, maka asumsi independensi lokal juga terpenuhi. Dalam studi ini, asumsi unidimensional sudah terpenuhi sehingga uji independensi lokal juga sudah terpenuhi. Asumsi berikutnya adalah invariansi parameter yang terdiri dari parameter butir dan kemampuan. Analisis invariansi parameter butir dan kemampuan responden menggunakan model logistik 3 parameter (3-PL) yang memuat koefisien tingkat kesukaran butir soal, daya beda, dan juga tebakan semu (*pseudo-guessing*). Untuk melihat invariansi parameter butir soal, responden dapat dibagi berdasarkan jenis kelamin, responden dengan nomor tes bernomor ganjil dan genap. Sedangkan untuk uji invariansi parameter kemampuan, instrumen dibagi menjadi 2 instrumen untuk butir ganjil dan butir genap.

Himpunan pasangan tingkat kesukaran untuk masing – masing butir pada setiap respon diamati di persekitaran garis dengan gradien sama dengan 1, atau persamaan garis $y = x$. Jika pasangan tingkat kesukaran untuk setiap butir soal ada mendekati garis tersebut, maka untuk parameter butir dikatakan bersifat invarian. Berikut hasil *plot* analisis parameter butir soal untuk model logistik 3 Parameter (3-PL) dengan responden pria dan wanita.

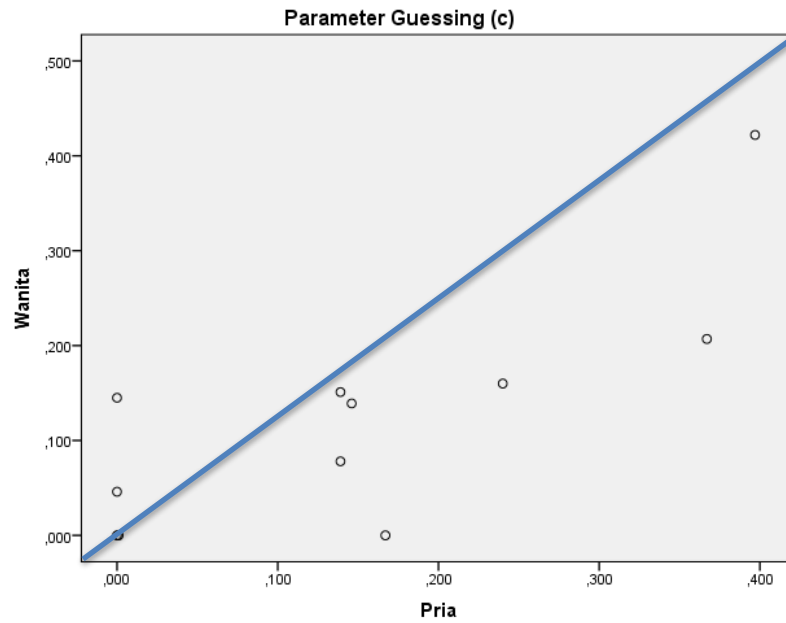


Gambar 2. Invariansi Parameter Daya Pembeda

Gambar 2 menunjukkan bahwa masing-masing titik berada relatif dekat dengan garis kemiringan 1. Hal ini menunjukkan bahwa tidak terjadi variasi parameter daya pembeda hasil estimasi pada kelompok pria dan kelompok wanita.



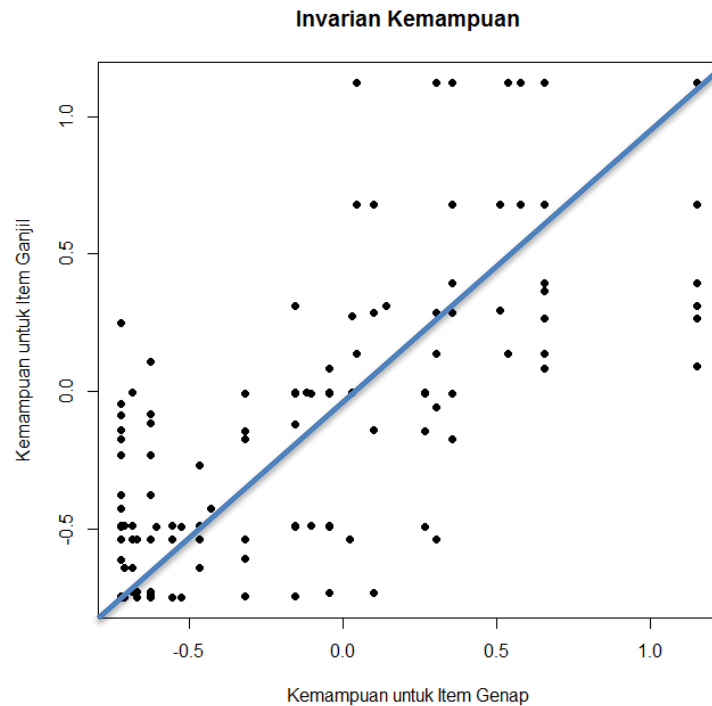
Gambar 3. Invariansi Parameter Tingkat Kesukaran (b)



Gambar 4. Invariansi Parameter Guessing (c)

Berdasarkan ketiga analisis ini dapat dibuktikan bahwa tidak terjadi variasi parameter a , b dan c (Gambar 4) hasil estimasi pada kelompok separuh pertama (Pria) dan kelompok separuh kedua (Wanita). Dengan kata lain, invariansi parameter butir terpenuhi. Menggunakan Software R, Invariansi parameter kemampuan dapat dibuktikan dengan mengestimasi parameter kemampuan menggunakan butir nomor ganjil dan genap. Selanjutnya dibuat diagram pencar dan dibandingkan kedekatannya dengan garis kemiringan 1 (Retnawati, [2014](#)).

Gambar 5 menunjukkan hasil *plot* atau diagram pencar uji invarians parameter kemampuan responden dengan menggunakan *Maximum A Posteriori* (MAP) dengan masing-masing butir soal ganjil dan genap sebagai pasangan titik-titik berada relatif dekat dengan garis kemiringan 1. Hal ini memberikan informasi bahwa tidak terjadi variasi parameter antara estimasi parameter kemampuan pada butir soal ganjil dan genap yang berarti syarat uji invariansi parameter kemampuan terpenuhi. Berdasarkan hasil keseluruhan analisis invariansi parameter butir dan kemampuan, maka diperoleh hasil bahwa invariansi parameter terbukti.



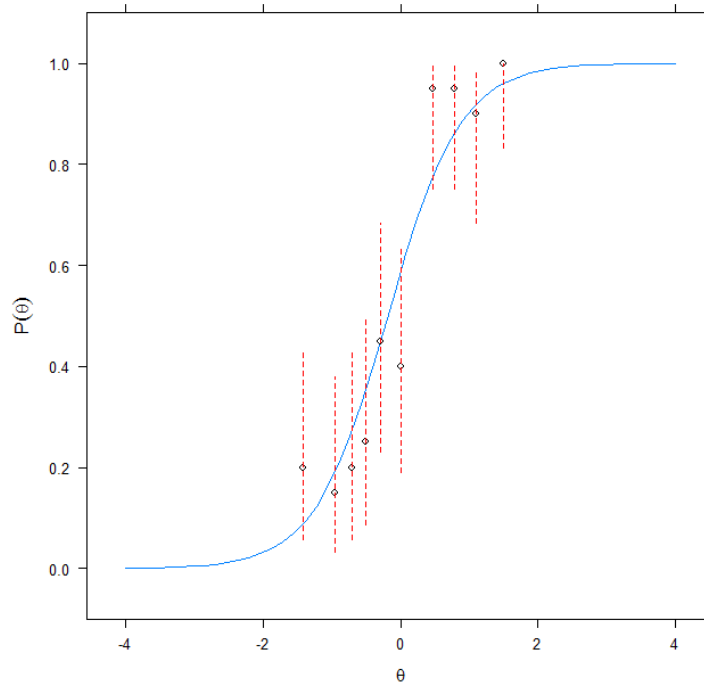
Gambar 5. Invariansi Parameter kemampuan dengan Estimasi butir Ganjil dan Genap

Uji kecocokan model teori respon butir dapat ditinjau berdasarkan identifikasi nilai khi-kuadrat untuk setiap butir soal dengan menggunakan model 1-PL, 2-PL, dan 3-PL. Hasil analisis kecocokan model 1-PL, 2-PL, dan 3-PL diperoleh dengan membandingkan nilai khi-kuadrat hitung yang diperoleh dengan menggunakan *software R* dibandingkan dengan nilai khi-kuadrat tabel untuk 11 butir soal. Hasil analisis lengkap kecocokan model parameter logistik untuk setiap butir soal diberikan dalam Tabel 4 berikut.

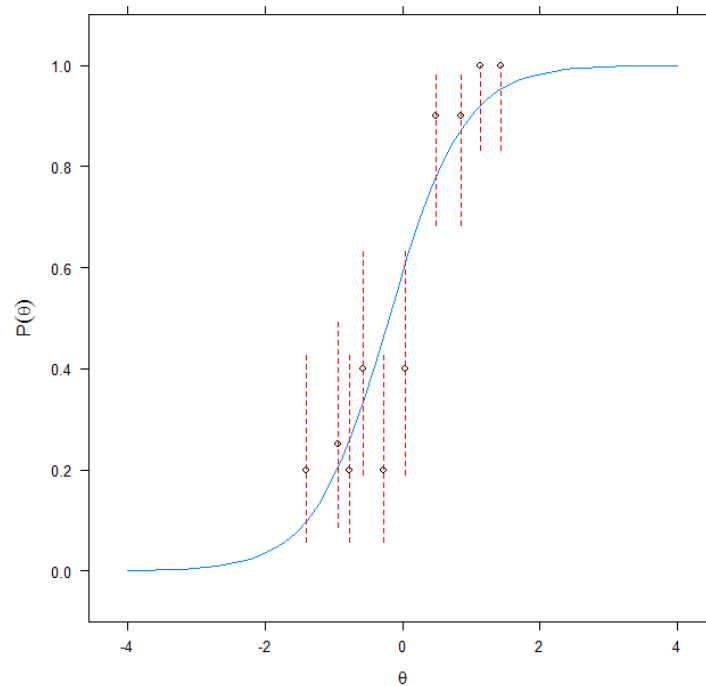
Tabel 4. Hasil Analisis Kecocokan Model Parameter Logistik pada Perangkat Tes Kemampuan Diagnosa Matematika Dasar

Item	Analisis Kecocokan Model		
	1-PL	2-PL	3-PL
Item1	-	-	√
Item2	-	√	√
Item3	-	-	-
Item5	-	√	√
Item6	√	√	√
Item7	√	√	√
Item8	-	√	√
Item9	-	√	√
Item10	-	√	√
Item11	-	-	√
Item12	√	√	√
Jumlah Item yang cocok	3	8	10

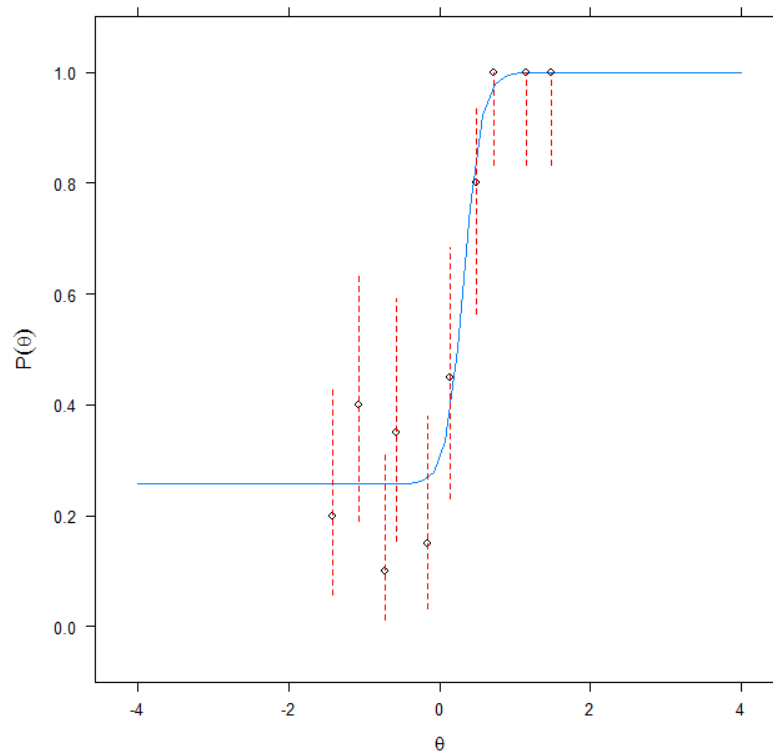
Tabel 4 memberikan informasi mengenai kecocokan model parameter logistik untuk 11 butir soal. Untuk model 1-PL, analisis kecocokan model hanya terdiri dari 3 butir soal, sedangkan untuk model 2-PL dan 3-PL terdapat 8 dan 10 butir soal yang cocok. Dengan demikian berdasarkan analisis kecocokan model, maka instrumen tes diagnosa kemampuan matematika dasar di atas mempunyai kecocokan dengan model logistik 3-PL dimana butir soal yang cocok sebanyak 10 butir. Selanjutnya, metode penentuan kecocokan model dengan menggunakan plot kurva karakteristik yang digambar dengan bantuan program R. Dengan plot ini, dapat diketahui seberapa tepat distribusi data dibandingkan dengan modelnya. Sebagai contoh pada butir nomor 12 seperti yang disajikan dalam Gambar 6. Pada Gambar 6, kurva karakteristik butir 12 model 1-PL, distribusi data cukup banyak yang letaknya lebih jauh dibandingkan dengan model 2-PL (Gambar 7) dan 3-PL (Gambar 8). Pada model 3-PL, distribusi data lebih mendekati dibandingkan 2-PL.



Gambar 6. Plot Kurva Karakteristik Item 12 Model 1-PL



Gambar 7. Plot Kurva Karakteristik Item 12 Model 2-PL



Gambar 8. Plot Kurva Karakteristik Item 12 Model 3-PL

Berdasarkan perbandingan pada ketiga plot kurva karakteristik di atas dapat diperoleh kesimpulan bahwa model 3-PL merupakan model yang paling baik untuk item nomor 12 dibandingkan model 2-PL dan 1-PL. Dengan prosedur yang serupa dapat dicari model yang sesuai untuk butir lainnya.. Hasil

analisis estimasi parameter instrumen tes diagnosa untuk model 3-PL diberikan pada Tabel 5 berikut.

Tabel 5. Hasil Analisis Identifikasi Kualitas Butir Model 3-PL

Item	Daya Beda (<i>a</i>)	Tingkat Kesukaran (<i>b</i>)	Tebakan Semu (<i>c</i>)	Kriteria
Item1	1,134	-0.772	0.000	Baik
Item2	21, 161	-0.088	0.018	Kurang Baik ($a > 2$)
Item3	2,327	-0.918	0.000	Kurang Baik ($a > 2$)
Item5	4,351	-0.053	0.079	Kurang Baik ($a > 2$)
Item6	4,962	0.277	0.196	Kurang Baik ($a > 2$)
Item7	5,898	0.143	0.117	Kurang Baik ($a > 2$)
Item8	1,549	0.137	0.094	Baik
Item9	14,810	0.798	0.387	Kurang Baik ($a > 2$) dan $c > 0,2$
Item10	3,877	0.901	0.135	Kurang Baik ($a > 2$)
Item11	23,485	0.578	0.152	Kurang Baik ($a > 2$)
Item12	16,670	0.553	0.272	Kurang Baik ($a > 2$) dan $c > 0,2$

Tabel 5 menunjukkan hasil analisis identifikasi kualitas untuk setiap butir perangkat tes diagnosa kemampuan matematika. Sebanyak 11 butir dengan kriteria butir soal baik dan kurang baik. Kriteria butir soal berdasarkan analisis kecocokan model dengan tingkat daya pembeda butir adalah 0.0 sampai 2.0, tingkat kesukaran butir soal, yakni -2.0 sampai 2.0, atau tingkat tebakan semu kurang dari 0.2, maka butir soal mempunyai kriteria baik. Sedangkan butir soal dengan kriteria kurang baik mempunyai tingkat daya pembeda butir adalah kurang dari 0.0 dan lebih dari 2.0, tingkat kesukaran butir soal, yakni kurang dari -2.0 dan lebih dari 2.0, atau tingkat tebakan semu lebih dari 0.2. Hasil analisis di atas menunjukkan hanya terdapat 2 butir soal yang memiliki kriteria baik sedangkan lainnya sebanyak 9 butir soal dengan kriteria kurang baik.

Menurut Mardapi (1998) terdapat tiga parameter butir yang diestimasi, yaitu tingkat kesukaran, daya pembeda dan tingkat tebakan. Tingkat kesukaran merupakan proporsi peserta dalam menjawab benar butir soal yang bisa diperoleh dengan melakukan analisis jumlah peserta tes yang menjawab benar butir soal dibandingkan dengan banyaknya peserta tes yang menjawab butir soal tersebut. Kemudian, daya pembeda untuk setiap butir soal memberikan informasi terkait butir soal yang dapat membedakan kemampuan peserta tes. Secara definisi, daya pembeda adalah korelasi antara skor butir tes dengan skor total dan disebut sebagai korelasi poin biserial. Ebel & Fresbie (1986) menyatakan bahwa korelasi poin biserial menunjukkan

korelasi antara skor butir tes dan skor total untuk setiap testi. Biserial dengan nilai positif dan tinggi menunjukkan kecenderungan testi yang mempunyai skor tinggi untuk menjawab benar dan testi yang berskor rendah memberikan jawaban yang salah. Sebaliknya, untuk nilai biserial yang negatif memberikan informasi mengenai kemampuan testi dengan skor tinggi memberikan jawaban salah dalam menjawab item, sedangkan testi dengan skor rendah benar dalam menjawab butir tersebut. Dalam melakukan analisis butir untuk memilih butir yang baik, maka butir dengan poin biserial yang negatif sebaiknya dikeluarkan dari model. Analisis butir klasik ini digunakan sebagai tahap awal dalam menentukan butir atau item yang cocok dengan model logistik parameter dengan melakukan analisis korelasi biserial. Analisisnya dengan mengidentifikasi nilai atau koefisien korelasi poin biserial untuk setiap butir yang bernilai negatif dan mengeliminasi butir tersebut dalam perangkat tes yang digunakan untuk analisis dan uji prasyarat model teori respon butir (IRT).

Instrumen atau perangkat tes diagnosa dalam penelitian ini terdiri dari 12 butir soal diagnosa kemampuan penguasaan materi eksponen dengan responden terdiri dari mahasiswa tingkat 1 sampai dengan 4 sebanyak 200 orang. Berdasarkan hasil analisis dengan menggunakan *software R* diperoleh 1 butir soal yang memiliki nilai korelasi poin biserial negatif sehingga butir soal tersebut tidak digunakan untuk analisis lebih lanjut. Butir soal tersebut butir soal nomor 4 dengan nilai korelasi poin biserial -0.1963 . Nilai korelasi poin biserial negatif tersebut menunjukkan bahwa mahasiswa dengan kemampuan tinggi memberikan respon salah dalam menjawab sedangkan untuk calon mahasiswa dengan kemampuan rendah memberikan respon benar dalam menjawab butir soal tersebut.

Menurut Embretson & Reise (2000), Hambleton & Swaminathan (1985:16), dan Hambleton, Swaminathan, & Rogers (1991:9) masing-masing menyatakan bahwa uji asumsi yang menjadi prasyarat teori respon butir adalah unidimensi, independensi lokal, dan invariansi parameter. Asumsi unidimensi dipenuhi, jika *plot* dari nilai-nilai Eigen berikutnya menjadi identik, kecuali untuk *plot* nilai Eigen pertama yang jauh lebih besar dari nilai-nilai Eigen lain. Menurut Retnawati (2017) berdasarkan analisis faktor diperoleh matrik varians-kovarians, dan selanjutnya melakukan penghitungan nilai Eigen yang digunakan untuk menghitung persentase varians yang terjelaskan dan juga sekaligus menggambarkan *scree-plot* yang menunjukkan tingkat kecuraman. Dalam penelitian ini, analisis uji unidimensi pada perangkat tes diagnosa menggunakan analisis faktor dengan *software SPSS*. Berdasarkan hasil analisis faktor diperoleh bahwa respon

mahasiswa terhadap perangkat tes diagnosa memuat 2 nilai eigen yang lebih besar dari 1 dimana terdapat 2 faktor atau sebanyak 53,733% varians yang dapat dijelaskan dalam perangkat tes tersebut dengan faktor pertama merupakan faktor dominan. Faktor dominan tersebut dapat diidentifikasi dengan melihat curaman berdasarkan *scree-plot* hasil analisis menggunakan *software* SPSS pada Gambar 1. Berdasarkan hasil analisis faktor dengan melihat nilai Eigen matriks varians kovarians dan *scree-plot* menunjukkan bahwa perangkat tes diagnosa kemampuan matematika dasar memenuhi syarat unidimensi.

Asumsi kedua yakni independensi lokal dideskripsikan sebagai respons terhadap suatu butir yang tidak dipengaruhi oleh respons lainnya. Menurut Retnawati (2014:3) uji asumsi independensi lokal dipenuhi untuk peluang dari pola jawaban setiap responden adalah hasil kali dari peluang jawaban responden untuk setiap butir soal. Kemudian, menurut Syarifah (2007) independensi lokal mensyaratkan bahwa untuk kemampuan θ yang tetap, setiap butir dalam instrumen tes tidak berkorelasi yang berarti butir-butir soal yang digunakan dalam penelitian tidak berkorelasi dalam kelompok untuk kemampuan θ bervariasi. Dalam hal ini, independensi lokal secara otomatis mengikuti keunidimensian. Dalam penelitian ini, uji independensi lokal berdasarkan syarat untuk setiap butir dalam instrumen tes diagnosa kemampuan matematika dasar tidak berkorelasi dalam kelompok tes untuk kemampuan θ bervariasi. Dalam hal ini, independensi lokal secara otomatis mengikuti keunidimensian. Oleh karena perangkat tes diagnosa memenuhi unidimensi berarti juga memenuhi syarat independensi lokal.

Uji asumsi model teori respon butir berikutnya adalah uji invariansi parameter. Menurut Retnawati (2014:3) invariansi parameter dapat didefinisikan sebagai keterkaitan karakteristik butir soal yang tidak mempunyai korelasi terhadap distribusi parameter kemampuan peserta tes dan juga parameter dengan karakteristik kemampuan peserta tes juga tidak berkorelasi dengan butir soal. Dalam hal ini, parameter butir tes tidak berbeda terhadap kelompok peserta tes dengan kemampuan yang bervariasi dan kemampuan responden tetap untuk butir soal dengan tingkat kesukaran yang berbeda. Menurut Hambleton, Swaminathan, & Rogers (1991:18), invariansi parameter butir dapat diketahui dengan memberikan tes kepada kelompok peserta dimana jika hasil estimasi parameter butir tetap untuk pengujian tes kepada kelompok peserta dengan kemampuan yang bervariasi, maka syarat invariansi parameter butir terpenuhi. Sedangkan untuk invariansi parameter kemampuan dapat diidentifikasi untuk perangkat tes yang berbeda memiliki tingkat kesukaran yang berbeda diujikan pada sekelompok peserta tes

dimana syarat invarians parameter kemampuan terpenuhi, jika hasil estimasi parameter kemampuan peserta tes tetap walaupun tes yang dikerjakan berbeda tingkat kesulitannya. Selanjutnya menurut Retnawati (2014:7), untuk invariansi parameter butir dapat juga dilakukan dengan memilih kemampuan responden berdasarkan jenis kelamin, tempat tinggal, status sosial ekonomi, dan lain-lain. Kemudian, hasil respon kelompok peserta tes terhadap perangkat tes dilakukan estimasi parameter butir diperoleh daya pembeda, tingkat kesukaran, dan tingkat tebakan semu. Masing-masing hasil estimasi dari pasangan parameter butir tersebut dikorelasikan dan jika korelasinya tinggi, maka asumsi invarians parameter butir terpenuhi. Pengujian invariansi parameter kemampuan juga dapat dilakukan dengan mengelompokkan butir soal yang bernomor ganjil dan genap. Kemudian dilakukan estimasi parameter kemampuan dengan menggunakan beberapa metode estimasi marginal yakni *Maximum Likelihood* (ML), *Maximum A Posteriori* (MAP), dan *Expected A Posteriori* (EAP) (Retnawati, 2015). Jika kemampuan responden berdasarkan respon terhadap butir ganjil dan genap memiliki korelasi tinggi, maka uji invariansi parameter kemampuan terpenuhi.

Uji invariansi parameter terdiri dari invariansi parameter butir dan kemampuan. Invariansi parameter butir dilakukan dengan membagi responden menjadi kelompok responden pria dan wanita dengan masing-masing responden sebanyak 100 mahasiswa. Kemudian, hasil respon terhadap perangkat tes dilakukan estimasi parameter butir, yakni parameter daya pembeda, tingkat kesukaran, dan tingkat tebakan semu. Hasil estimasi masing-masing parameter tersebut dilakukan *plot* berdasarkan kedua kelompok responden pada garis lurus dengan kemiringan sebesar 1. Gambar 2, 3, dan 4 menunjukkan bahwa beberapa pasangan titik yang merupakan respon terhadap 2 kelompok responden mendekati garis lurus $y = x$ yang berarti mempunyai korelasi tinggi. Dengan demikian, perangkat tes diagnosa memenuhi uji invariansi parameter butir. Uji invariansi berikutnya adalah invariansi parameter kemampuan. Invariansi parameter kemampuan dilakukan dengan membagi butir menjadi butir soal ganjil sebanyak 6 dan butir genap sebanyak 5 soal dengan responden sebanyak 200 mahasiswa. Hasil respon calon mahasiswa terhadap butir soal ganjil dan genap dianalisis menggunakan estimasi MAP kemudian dilakukan *plot* berdasarkan respon terhadap butir ganjil dan genap pada garis lurus dengan kemiringan 1. Pada Gambar 5 terlihat bahwa pasangan titik yang merupakan respon calon mahasiswa terhadap kedua jenis butir soal mendekati garis lurus $y = x$ yang berarti mempunyai korelasi tinggi. Dengan demikian, perangkat tes diagnosa kemampuan matematika dasar memenuhi uji invariansi parameter

kemampuan. Oleh karena uji invariansi parameter butir dan kemampuan dipenuhi maka perangkat tes diagnosa kemampuan matematika dasar memenuhi uji invariansi parameter.

Analisis berikutnya yakni uji kecocokan model teori respon butir pada perangkat tes diagnosa. Menurut Hambleton, Swaminathan, & Rogers (1991) model teori respon butir terdiri dari model logistik 1 parameter (1-PL), 2-PL, dan 3-PL. Banyaknya parameter disesuaikan dalam model logistik tersebut, seperti untuk model 1-PL hanya terdiri parameter tingkat kesukaran saja, dan untuk 2-PL memuat parameter daya beda dan tingkat kesukaran sedangkan untuk model 3-PL berupa daya beda, tingkat kesukaran, dan tebakan semu. Uji kecocokan model terhadap data respon dalam perangkat tes merupakan tolak ukur yang digunakan untuk memilih model analisis yang dikenakan pada data tersebut. Analisis kecocokan model ini merupakan bagian penting dalam melakukan estimasi kemampuan individu. Pemilihan model analisis yang tepat memberikan informasi tingkat kesalahan kecil dalam melakukan estimasi kemampuan per individu. Akan tetapi, menurut Wiberg (2004) pada dasarnya model yang terbaik tidak secara sempurna mempunyai kecocokan dengan data. Analisis penghitungan khi-kuadrat hitung untuk setiap butir menggunakan *software R* sedangkan nilai khi-kuadrat tabel menggunakan derajat kebebasan sebesar 0,05. Hasil analisis perbandingan nilai khi-kuadrat hitung dan nilai khi-kuadrat tabel diberikan pada Tabel 3. yang menunjukkan bahwa banyaknya butir soal yang cocok dengan model logistik 1 parameter (1-PL) sebanyak 3 butir soal, sedangkan untuk model 2-PL dan 3-PL masing-masing terdapat 8 dan 11 butir soal. Dengan demikian maka instrumen tes diagnosa di atas mempunyai kecocokan dengan model logistik 3-PL dengan banyaknya butir soal yang cocok sebanyak 11 butir soal.

Analisis kecocokan model berikutnya menggunakan pendekatan *plot* kurva karakteristik butir (ICC). Menurut Retnawati (2014) *plot* kurva karakteristik butir memberikan informasi butir terkait ketepatan distribusi data dengan model logistik parameter yang digunakan. Hal ini berarti, jika terdapat perbandingan model logistik parameter perilaku berdasarkan *plot* ICC, maka model logistik parameter yang cocok adalah model dengan perilaku distribusi data lebih mendekati kurva nonlinear yang diberikan masing-masing model logistik tersebut. Dalam penelitian ini, diberikan *plot* kurva karakteristik butir 12 untuk melihat perilaku distribusi data mendekati kurva nonlinear model logistik model 3-PL. Hasil *plot* kurva karakteristik butir 12 diberikan pada Gambar 6, 7, dan 8 yang menunjukkan bahwa perilaku distribusi data untuk model 3-PL relatif lebih dekat dibandingkan dengan model 2-PL dan 1-PL. Hal ini menunjukkan model 3-PL berdasarkan pendekatan *plot* ICC

merupakan model yang paling cocok digunakan dibandingkan dengan model logistik lainnya, yakni model 1-PL dan 2-PL. Selanjutnya, berdasarkan analisis kecocokan model teori respon butir terhadap perangkat tes diagnosa diperoleh model 3-PL dengan tingkat kecocokan butir yang paling banyak, sehingga perlu untuk identifikasi butir soal tes dengan kriteria baik berdasarkan kriteria teori respon butir. Selain itu, informasi yang bisa diperoleh dengan analisis identifikasi butir soal tersebut sebagai bahan pertimbangan untuk pemilihan butir soal dalam perangkat tes tersebut.

Menurut Hulin, Drasgow, & Parsons (1983) identifikasi butir dengan kriteria baik dan kurang baik berdasarkan indeks parameter daya pembeda (a), tingkat kesukaran (b), dan indeks tebakan semu (c). Indeks daya pembeda butir diantara 0.0 dan 2.0, tingkat kesukaran pada interval -2.0 dan 2.0, dan indeks tebakan semu lebih kecil dari 0.2, maka butir soal tersebut memiliki kriteria baik. Sebaliknya, butir soal tergolong dalam kriteria kurang baik, jika memiliki indeks daya pembeda butir lebih kecil dari 0.0 dan lebih besar 2.0, tingkat kesukaran lebih kecil -2.0 dan lebih besar dari 2.0, dan indeks tebakan semu lebih besar 0.2. Selain itu, menurut Hambleton & Swaminathan (1985) butir yang baik dengan indeks daya beda terletak pada interval 0.0 dan 2.0, sedangkan tingkat kesukaran menurut Allen & Yen (1979) berada pada interval 0.3 dan 0.7 dan untuk indeks tebakan semu tidak lebih dari nilai $1/k$ dengan k menyatakan banyaknya pilihan jawaban pada perangkat tes (Hulin, Drasgow, & Parsons, 1983). Hasil analisis identifikasi butir berdasarkan estimasi parameter butir model 3-PL diberikan pada Tabel 5 menunjukkan bahwa terdapat 2 butir soal, yakni butir 1 dan 8 dengan kriteria baik dimana indeks daya pembeda butir berada pada interval 0.0 dan 2.0, tingkat kesukaran diantara -2.0 dan 2.0, dan indeks tebakan semu lebih kecil dari 0.2. Sedangkan 9 butir lainnya merupakan kriteria butir-butir kurang baik dengan 7 butir soal mempunyai daya pembeda lebih dari 2.0, yakni butir 2, 3, 5, 6, 7, 10, dan 11. Kemudian, butir soal dengan indeks daya beda soal lebih besar daripada 2.0 dan indeks tebakan semu yang lebih besar daripada 0.2 sebanyak 2 butir soal, yakni butir 9 dan 12.

Jika meninjau keseluruhan parameter butir nampak hanya 2 butir soal yang memenuhi kriteria baik dari keseluruhan parameter butir. Meskipun begitu, jika ditinjau dari tingkat kesulitan butir maka keseluruhan butir memiliki tingkat kesulitan yang baik. Parameter tingkat kesulitan butir umumnya yang dijadikan asumsi penyusunan soal, bukan parameter yang lain seperti daya pembeda atau tebakan semu. Merujuk pada referensi tersebut, maka dapat dikatakan seluruh butir memiliki karakteristik butir yang baik.

Simpulan

Berdasarkan nilai korelasi poin biserial dari 12 butir soal diagnosa, diperoleh informasi bahwa nilai korelasi poin biserial butir nomor 4 adalah negatif, sehingga butir tersebut tidak digunakan dalam analisis parameter dan kecocokan model. Selanjutnya berdasarkan analisis dengan metode statistik dan grafik, model teori respons butir yang cocok dengan soal diagnosa kemampuan matematika dasar untuk materi eksponen adalah model 3-PL. Dari nilai estimasi terhadap masing-masing parameter diperoleh tingkat kesukaran item berada dalam rentang $-0,918$ sampai $0,901$. Daya pembeda berada dalam rentang $1,134$ sampai $21,161$ serta tebakan semu ada di antara nilai 0 sampai dengan $0,387$. Analisis lanjutan, dengan meninjau nilai parameter masing-masing butir ditemukan bahwa dari 11 butir yang dianalisis sekitar 18,2% (dua butir soal) yang memenuhi kategori kualitas butir yang baik jika dilihat dari parameter daya pembeda. Meskipun demikian, jika ditinjau dari tingkat kesulitan butir terbukti bahwa keseluruhan butir memiliki tingkat kesulitan yang baik, sehingga dapat disimpulkan bahwa seluruh butir memiliki karakteristik butir yang baik.

Daftar Pustaka

- Allen, M.J., & Yen, W.M. (1979). *Introduction to Measurement Theory*. California: Wadsworth, Inc.
- De Mars, C.E. (2010). *Item Response Theory*. New York: Oxford University Press.
- Ebel, R.L. & Frisbie, D.A. (1986). *Essentials of Educational Measurement*. New Jersey: Prentice-Hall, Inc.
- Embretson, S., & Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory*. Boston, MA: Kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. London: Sage.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item Response Theory: Application to Psychological Measurement*. Homewood, IL: Irwin.
- Leighton, J. P., & Gierl, M. J. (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge: Cambridge University Press.
- Levine, M. D., Lindsay, R. L., & Reed, M. S. (1992). The Wrath of Math: Deficiencies of Mathematical Mastery in The School Child. *Pediatric Clinics of North America*, 39(3), 525–536. [https://doi.org/10.1016/S0031-3955\(16\)38342-0](https://doi.org/10.1016/S0031-3955(16)38342-0)
- Mardapi, D. (1998). Analisis Butir dengan Teori Klasik dan Teori Respon Butir. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran* 28(1), 25–34. <https://doi.org/10.21831/jk.v28i1.7244>
- Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Nuha Medika.
- Retnawati, H. (2015). Perbandingan Estimasi Kemampuan Laten antara Metode Maksimum Likelihood dan Metode Bayes. *Jurnal Penelitian dan Evaluasi Pendidikan* 19(2), 146–155. <https://doi.org/10.21831/pep.v19i2.5575>

- Retnawati, H. (2017). *Validitas, Reliabilitas & Karakteristik Butir: Panduan untuk Peneliti, Mahasiswa, dan Psikometri*. Yogyakarta: Parama Publishing.
- Stone, C. A., & Zhang, B. (2003). Assessing Goodness of Fit of Item Response Theory Models: A Comparison of Traditional and Alternative Procedures. *Journal of Educational Measurement*, 40(4), 331-352. <https://doi.org/10.1111/j.1745-3984.2003.tb01150.x>
- Syarifah. (2007). Persyaratan Analisis Instrumen sebagai Prasyarat Ketepatan Hasil Analisis dalam Penelitian Pendidikan. *Jurnal Penelitian MIPA*, 1(1), 12 – 19.
- Tall, D., & Razali, M. R. (1993). Diagnosing Students' Difficulties in Learning Mathematics. *International Journal of Mathematical Education in Science and Technology*, 24(2), 209–222. <https://doi.org/10.1080/0020739930240206>
- Wiberg, M. (2004). Classical Test Theory vs. Item Response Theory. *Umea*, 10(5), 1–27.