



## PERBANDINGAN KECOCOKAN MODEL ANALISIS TEORI RESPONS BUTIR PADA DATA POLITOMI

Ndaru Wicaksono Aji<sup>a\*</sup>, Heri Retnawati<sup>a</sup>

<sup>a</sup>Program Studi Statistika, Universitas Negeri Yogyakarta

\*ndaruwicaksono89@gmail.com

**Abstrak.** Perbandingan kecocokan model analisis teori respons butir politomi dilakukan supaya diperoleh model yang paling baik dan cocok. Penelitian ini bertujuan untuk mengetahui kecocokan pada model *Graded Response Model* (GRM), *Partial Credit Model* (PCM), dan *Generalized Partial Credit Model* (GPCM) pada data politomi. Penelitian ini merupakan penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Perbandingan kecocokan model dengan uji *Yen's Q1* dan melihat nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC). Data yang digunakan yaitu 851 data jawaban peserta tes terhadap 50 soal *reading* pilihan ganda yang mengukur kemampuan berbahasa Inggris tahun 2021. Model dengan butir cocok paling banyak dan nilai AIC dan BIC terkecil dipilih paling cocok. Hasil penelitian ini adalah 1) model GRM sebanyak 5 butir yang cocok pada butir 1, 3, 4, 5, dan 6. 2) model PCM terdapat 4 butir yang cocok pada butir 3, 4, 5, dan 6. 3) model GPCM sebanyak 5 butir yang cocok pada butir 1, 3, 4, 5, dan 6. 4) model yang paling baik dan paling cocok untuk digunakan adalah GPCM karena berdasarkan uji kecocokan butir GPCM dan GRM mempunyai jumlah butir cocok sama, tetapi untuk nilai AIC dan BIC pada GPCM yang paling kecil sehingga GPCM terpilih sebagai model terbaik.

**Kata kunci:** tes kemampuan berbahasa Inggris, teori respons butir, politomi, reading, *Yen's Q1*

**Abstract.** Comparison of fit of the polytomous item response theory analysis model was carried out in order to obtain the best and most suitable model for scoring on the test. This study aims to determine the suitability of the *Graded Response Model* (GRM), *Partial Credit Model* (PCM), dan *Generalized Partial Credit Model* (GPCM) in the polytomous data. Comparison of models fit with *Yen's Q1* test and looked at the *Akaike Information Criterion* (AIC) and *Bayesian Information Criterion* (BIC) values. The data used are 851 test participants' answers to 50 multiple-choice reading questions that measure English language skills in 2021. The model with the most suitable items and the smallest AIC and BIC values is selected the most suitable model. The results of this study are 1) in the GRM model as many as 5 items that are fit in item 1, 3, 4, 5, and 6. 2) in the PCM model there are 4 items that are fit in item 3, 4, 5, and 6. 3) in the GPCM model as many as 5 items are fit in item 1, 3, 4, 5, and 6. 4) the best and most suitable model to be used is the GPCM model because based on the item compatibility test GPCM and GRM had the same number of matched items, but the AIC and BIC values in the GPCM were the smallest so that GPCM was chosen as the best model.

**Keywords:** English proficiency test, item response theory, polytomous, reading, *Yen's Q1*



## PENDAHULUAN

Tes sering dilakukan untuk seleksi atau sebuah pengujian terhadap kemampuan seseorang. Tes merupakan suatu langkah atau prosedur yang digunakan dalam pengukuran dan penilaian dalam suatu bidang dalam pendidikan (Kadir, 2015). Pada umumnya, tes berisi kumpulan soal yang digunakan untuk menguji seseorang sehingga menghasilkan gambaran tentang kemampuan atau kepribadian seseorang (Adiyo, 2014). Hasil dari suatu tes pada mulanya hanya berupa soal tersebut terjawab benar atau salah. Hasil tersebut dapat dikategorikan kedalam penskoran dikotomi dengan mengkonversi nilai hasil tes kedalam nilai 0 dan 1.

Seiring perkembangan zaman, muncul hasil tes dengan penskoran yang mempunyai level atau bobot nilai. Hasil skor yang mempunyai beberapa level tersebut disebut penskoran politomi. Penggunaan penskoran politomi dianggap lebih baik dalam penentuan hasil atau penentuan keputusan. Hal ini sesuai dengan penelitian yang dilakukan oleh Sepuzaman et al. (2021) yang melakukan estimasi kemampuan dengan teori respons butir penskoran dikotomi dan politomi dengan hasil analisis yang lebih baik dalam memetakan kemampuan peserta tes dengan penskoran dengan politomus. Pada penskoran politomi setiap soal mempunyai tingkat nilai yang berbeda tergantung dengan tingkat kesulitannya.

Analisis terhadap butir-butir soal tersebut dapat menggunakan teori respons butir. Teori respons butir merupakan salah satu teori pengukuran modern yang dapat digunakan untuk menganalisis butir soal. Dalam hal ini, probabilitas peserta tes yang mempunyai kemampuan lebih tinggi akan lebih besar dibandingkan probabilitas peserta tes yang mempunyai kemampuan rendah (Retnawati, 2014). Analisis butir soal dengan teori respons butir pada data hasil yang mempunyai skor yang merupakan data politomi ordinal dapat menggunakan beberapa model analisis seperti *Graded Respons Model* (GRM), *Partial Credit Model* (PCM), dan *Generalized Partial Credit Model* (GPCM). Dari model-model yang ada perlu diketahui tingkat kecocokan pada model atau model yang paling baik. Untuk itu perlu dilakukan uji kecocokan model guna mengetahui model yang paling cocok dalam menguji analisis butir soal. Dengan diketahuinya model yang paling cocok untuk sebuah analisis respons butir sehingga akan menghasilkan keadaan dari kemampuan peserta tes yang sebenarnya dan yang paling tepat.

Model GRM mempunyai keunggulan dapat digunakan pada hasil penskoran maupun hasil skala likert. Tetapi dalam model GRM dalam penskorannya harus terurut. Model PCM mempunyai keunggulan tidak memerlukan ambang batas untuk mengikuti urutan yang sama dengan kategori respons (Desjardins & Bulut, 2018). Sedangkan model GPCM terdapat daya pembeda dan faktor skala sehingga setiap item dalam GPCM memiliki parameter diskriminasi butir yang unik (Desjardins & Bulut, 2018).

Penelitian mengenai pemodelan analisis data politomi telah banyak dilakukan. Safaruddin et al. (2012) meneliti tentang penskoran pada data politomi dengan *Partial Credit Model* (PCM) pada data nilai mahasiswa Fakultas Perikanan Unhas pada ujian Mid Semester mata kuliah Matematika Dasar tahun ajaran 2011/2012. Diperoleh hasil untuk butir soal X yang layak untuk uji tes adalah butir soal nomor 1 dan 7, dan untuk butir soal Y yang layak untuk uji tes adalah butir soal nomor 2, 7, 8, dan 12. Penelitian yang dilakukan oleh Adiyo (2014) mengenai pengaruh jumlah kategori, rentang threshold dan metode estimasi terhadap tingkat akurasi nilai tes pada beberapa model politomi. Penelitian tersebut diperoleh kesimpulan jika

aplikasi pada data empiris diduga model politomi GRM dan metode estimasi WLSMV atau MLR lebih menghasilkan nilai tes yang presisi.

Sehingga kesesuaian dari butir-butir soal yang diberikan mempunyai pengaruh terhadap hasil akhir dan kesimpulan dari suatu kemampuan yang diperoleh dari seorang peserta tes. Sangat penting adanya perangkat tes yang mempunyai kualitas baik. Perangkat tes yang berkualitas baik dapat dilihat dengan melakukan analisis terhadap butir-butir soal. Analisis ini dapat dilihat dengan melakukan pemodelan. Namun, belum diketahui model yang paling cocok dalam pemodelan tersebut. Oleh karena itu, perlu dilakukan perbandingan tingkat kecocokan model-model tersebut.

Pada penelitian ini, perbandingan tingkat kecocokan model analisis respons butir dengan data politomi hasil tes kemampuan berbahasa Inggris dengan model GRM, PCM, dan GPCM. Pemilihan ketiga model tersebut karena data yang digunakan merupakan data ordinal. Model politomi ordinal yang sering digunakan yaitu GRM, PCM, dan GPCM (Retnawati, 2014). Berbeda dari beberapa penelitian yang diutarakan pada paragraf sebelumnya, penelitian ini melakukan perbandingan terhadap model teori respons butir yang digunakan pada data politomi khususnya pada model untuk jenis data ordinal. Selain itu pada penelitian ini menggunakan data hasil tes kemampuan berbahasa Inggris.

Tes pengukur kemampuan berbahasa Inggris biasanya digunakan dalam melakukan uji standarisasi dari peserta tes dalam pemahaman dan kemampuan menggunakan bahasa Inggris. Terdapat beberapa tes pengukur kemampuan berbahasa Inggris yang ada. Diantara tes tersebut yaitu TOEFL (Test of English as a Foreign Language), ProTEFL (*Proficiency Test of English as a Foreign Language*), dan lain-lain. Tes pengukur kemampuan berbahasa Inggris mengukur peserta tes dalam bahasa Inggris yang meliputi keterampilan dalam mendengarkan (*Listening*), kemampuan tata bahasa atau menulis (*Grammar/writing*), dan membaca (*Reading*). Dalam keterampilan *reading* biasanya terdapat sebuah bacaan dengan berbagai topik yang nantinya digunakan untuk menjawab beberapa soal. Soal yang diberikan berupa pertanyaan mengenai informasi atau isi dari bacaan tersebut baik informasi secara *implisit* atau *eksplisit*.

## METODE

### Deskripsi Data

Pendekatan penelitian yang digunakan adalah penelitian deskriptif eksploratif dengan pendekatan kuantitatif. Pada penelitian ini, studi kasus menggunakan data 851 responden jawaban peserta tes terhadap soal *reading* pilihan ganda 50 soal yang mengukur kemampuan mahasiswa dalam berbahasa Inggris di salah satu universitas di Indonesia pada tahun 2021. Berikut contoh bacaan pada Gambar 1 dan soal *reading* pada Gambar 2 berikut.

#### Questions 1-5

A distinctively American architecture began with Frank Lloyd Wright, who had taken to heart the admonition that form should follow function and who thought of buildings not as separate architectural entities but as parts of an organic whole that included the land, the community, and the society. In a very real way the houses of colonial New England and some of the southern plantations had been functional, but Wright was the first architect to make functionalism the authoritative principle for public as well as for domestic buildings. As early as 1906 he built the Unity Temple in Oak Park, Illinois, the first of those churches that did so much to revolutionize ecclesiastical architecture in the United States. Thereafter he turned his genius to such miscellaneous structures as houses, schools, office buildings, and factories, among them the famous Larkin Building in Buffalo, New York, and the Johnson Wax Company building in Racine, Wisconsin.

(Sumber Gambar: <https://www.digitaleduka.com/artikel-ilmu-kunci-toefl/artikel/reading-comprehension/contoh-soal-pembahasan-reading-comprehension-tes-toefl>)

### Gambar 1. Contoh Bacaan *Reading*

- |  |  |
|--|--|
| <p>1. The phrase "taken to heart" in line 1 is closest in meaning to which of the following?</p> <p>(A) Taken seriously<br/>(B) Criticized<br/>(C) Memorized<br/>(D) Taken offence</p> <p>2. In what way did Wright's public buildings differ from most of those built by earlier architects?</p> <p>(A) They were built on a larger scale.<br/>(B) Their materials came from the southern United States.<br/>(C) They looked more like private homes.<br/>(D) Their designs were based on how they would be used.</p> <p>3. The author mentions the Unity Temple because, it</p> <p>(A) was Wright's first building<br/>(B) influenced the architecture of subsequent churches<br/>(C) demonstrated traditional ecclesiastical architecture<br/>(D) was the largest church Wright ever designed</p> | <p>4. The passage mentions that all of the following structures were built by Wright EXCEPT</p> <p>(A) factories<br/>(B) public buildings<br/>(C) offices<br/>(D) southern plantations</p> <p>5. Which of the following statements best reflects one of Frank Lloyd Wright's architectural principles?</p> <p>(A) Beautiful design is more important than utility.<br/>(B) Ecclesiastical architecture should be derived from traditional designs.<br/>(C) A building should fit into its surroundings.<br/>(D) The architecture of public buildings does not need to be revolutionary</p> |
|--|--|

(Sumber Gambar: <https://www.digitaleduka.com/artikel-ilmu-kunci-toefl/artikel/reading-comprehension/contoh-soal-pembahasan-reading-comprehension-tes-toefl>)

### Gambar 2. Contoh Soal *Reading*

Contoh dari Gambar 2 merupakan contoh soal *reading* dengan satu bacaan yang berhubungan dengan bacaan pada Gambar 1 tersebut.

Penskoran dalam soal *reading* dapat dilakukan dalam penskoran politomi. Soal-soal yang berhubungan dengan sebuah bacaan tersebut dijadikan satu. Penggabungan beberapa butir menjadi satu butir disebut model testlet. Menurut Widhiarso dan Suhapti (2018) testlet dalam penskoran politomi merupakan skor yang merepresentasikan kumpulan dari beberapa butir. Dengan kata lain setiap bacaan dianggap satu butir. Pembatasan masalah pada penelitian ini terletak pada perbandingan kecocokan model teori respons butir politomi pada data tes kemampuan berbahasa Inggris.

### Langkah Analisis Data

Analisis data yang dilakukan adalah sebagai berikut

1. Memisah hasil respons pada butir soal *Reading* dengan butir lainnya.
2. Menguji asumsi unidimensi, invariansi parameter, dan independensi lokal kelompok butir
  - a. Unidimensi

Asumsi unidimensi dapat dimaksudkan setiap butir soal hanya digunakan untuk mengukur satu kemampuan atau satu ciri peserta tes. Hal tersebut sesuai dengan Sudaryono (2011) yang berpendapat jika butir digunakan untuk mengukur lebih dari satu dimensi maka hasil jawaban butir tersebut gabungan dari lebih dari satu kemampuan peserta tes. Hal ini akan menyebabkan tidak diketahuinya pengaruh dari setiap kemampuannya dalam menjawab butir tes. Dalam pengujian asumsi unidimensi dapat menggunakan analisis faktor dengan melihat nilai eigen yang sudah diurutkan dan divisualisasikan dalam bentuk *screeplot*. Uji asumsi unidimensi terpenuhi jika terdapat patahan pada garis plot pada nilai eigen kedua atau dari nilai eigen pertama menuju nilai eigen kedua terjadi penurunan yang curam (DeMars, 2010).

- b. Independensi lokal

Analisis teori respons butir mempunyai syarat nilai suatu butir soal tidak bergantung oleh nilai pada butir soal yang lainnya. Hal tersebut dapat dilakukan pengujian dengan uji independensi lokal. Independensi lokal merupakan respons peserta tes terhadap pasangan butir-butir akan independen satu sama lain secara statistik ketika faktor-faktor yang mempengaruhi konstan (Retnawati, 2014). Dalam asumsi independensi lokal terpenuhi jika jawaban terhadap butir soal dari peserta tes tidak mempunyai pengaruh terhadap butir soal yang lainnya. Uji independensi lokal

dibuktikan dengan melihat korelasi antar butir pada data yang digunakan. Menurut Zainal, Effendi, & Matore (2020) independensi lokal terpenuhi jika korelasi antar butir kurang dari 0,7. Pada uji asumsi independensi lokal akan terpenuhi jika uji asumsi unidimensi juga terpenuhi (DeMars, 2010).

c. Invariansi parameter

Invariansi parameter mempunyai arti jika butir soal mempunyai karakteristik yang independen terhadap distribusi parameter kemampuan dari peserta tes dan parameter dari karakter peserta tes tidak berpengaruh dengan karakter butir dari soal (Hambleton, Swaminathan, & Rogers, 1991). Invariansi parameter kemampuan diuji menggunakan dua atau lebih perangkat tes yang mempunyai tingkatan kesulitan yang berbeda pada kelompok peserta tes. Invariansi parameter kemampuan terbukti jika hasil estimasi kemampuan peserta tes tidak berbeda untuk tes yang berbeda tingkatan kesulitannya. Selain itu, invariansi parameter butir dapat diselidiki dengan uji tes kelompok peserta yang berbeda. Jika diperoleh hasil estimasi parameter butir tidak berbeda meskipun diujikan pada kelompok yang berbeda tingkat kemampuannya maka invariansi parameter butir terbukti.

3. Melakukan pemodelan (GRM, PCM, dan GPCM) pada data dengan menggunakan pendekatan teori respon butir.

Pemodelan IRT mengukur respons dari peserta tes dalam level butir soalnya. Model-model IRT pada politomi sendiri dibagi kedalam model IRT nominal dan model IRT ordinal. Model-model IRT ordinal tersebut sebagai berikut.

a. *Graded Response Model (GRM)*

*Graded Respon Model (GRM)* merupakan pengembangan model 2 PL pada dikotomi. Model GRM umumnya digunakan pada data yang mempunyai hasil berjenjang atau terdapat tingkatan pada setiap butirnya. Dalam model GRM urutan penskoran harus terurut dan tidak boleh dibolak-balik. Model matematis hubungan parameter butir soal dengan kemampuan peserta untuk model GRM menurut Hambleton, Swaminathan, & Rogers (1991) dengan respons peserta tes dikategorikan dengan  $m + 1$  dari rendah ke tinggi dimisalkan  $k = 0, 1, 2, \dots$ . Nilai  $m$  merupakan banyaknya kategori dalam menjawab butir  $j$ . Persamaan model GRM adalah sebagai berikut.

$$P_{jk}(\theta) = P_{jk}^*(\theta) - P_{jk+1}^*(\theta) \quad (1)$$

dengan persamaan  $P_{jk}(\theta)$  sebagai berikut.

$$P_{jk}(\theta) = \frac{\exp [D a_j (\theta - b_{jk})]}{1 + \exp [D a_j (\theta - b_{jk})]} \quad (2)$$

dengan,

$P_{jk}^*(\theta)$  = probabilitas peserta tes dengan kemampuan  $\theta$  yang memperoleh nilai skor kategori  $k$  atau lebih pada butir  $j$ .

$P_{jk}(\theta)$  = probabilitas peserta tes dengan kemampuan  $\theta$  yang memperoleh nilai skor kategori  $k$  pada butir  $j$ .

$a_j$  = merupakan indeks daya beda pada butir  $j$ .

$\theta$  = merupakan kemampuan peserta tes,

$D$  = merupakan faktor skala sebesar 1,7,

$b_{jk}$  = indeks kesukaran kategori  $k$  pada butir  $j$ ,

Daya beda merupakan indeks yang membedakan kemampuan peserta kelompok atas dan kelompok bawah. Daya beda mempunyai persamaan sebagai berikut.

b. *Partial Credit Model (PCM)*

*Partial Credit Model (PCM)* merupakan salah satu pemodelan politomi pada IRT. PCM merupakan pengembangan dari model *Rasch* pada pemodelan dikotomi. model *Rasch* yang terdapat satu parameter dikembangkan dengan penjabaran parameter kedalam kategori. Pada model PCM diasumsikan jika setiap butir mempunyai nilai yang sama pada semua daya bedanya. Dalam model PCM dapat dikatakan jika dalam bagian-bagian atau setiap langkah terdapat skornya. Bentuk umum *Partial Credit Model* yaitu,

$$P_{jk}(\theta) = \frac{\exp \sum_{v=0}^k (\theta - b_{jk})}{\sum_{h=0}^m \exp \sum_{v=0}^h (\theta - b_{jk})} \quad (3)$$

$$k = 0, 1, 2, 3, \dots, m$$

Dengan,

$P_{jk}(\theta)$  = probabilitas peserta tes dengan kemampuan  $\theta$  yang memperoleh nilai skor kategori k pada butir j.

$\theta$  = merupakan kemampuan peserta tes,

$b_{jk}$  = indeks kesukaran kategori k pada butir j.

c. *Generalized Partial Credit Model (GPCM)*

*Generalized Partial Credit Model (GPCM)* mempunyai probabilitas penskoran kategori k pada butir. Model GPCM dalam menghitung estimasi kemampuan peserta memperhitungkan tingkat kesulitan pada setiap langkah-langkahnya. Model GPCM sendiri mirip dengan *Partial Credit Model (PCM)*. Tetapi pada model GPCM terdapat parameter daya pembeda ( $a$ ) dan faktor skala ( $D$ ) yang merupakan faktor skala yang telah ditetapkan sebesar 1,7. Model matematisnya sebagai berikut.

$$P_{jk}(\theta) = \frac{\exp \sum_{v=0}^h Z_{jh}(\theta)}{\sum_{h=0}^m \exp [\sum_{v=0}^h Z_{jh}(\theta)]}, k = 0, 1, 2, 3, \dots, m \quad (4)$$

Dengan  $Z_{jh}(\theta)$  persamaan sebagai berikut.

$$Z_{jh}(\theta) = D a_j (\theta - b_{jh}) = D a_j (\theta - b_j + d_h), b_{j0} = 0 \quad (5)$$

Dengan,

$P_{jk}(\theta)$  = probabilitas peserta tes dengan kemampuan  $\theta$  yang memperoleh nilai skor kategori k pada butir j.

$a_j$  = merupakan indeks daya beda pada butir j.

$\theta$  = merupakan kemampuan peserta tes,

$D$  = merupakan faktor skala sebesar 1,7,

$b_{jh}$  = indeks kesukaran pada kategori k butir j,

$b_j$  = indeks kesukaran pada lokasi butir j

$d_h$  = parameter pada kategori k.

4. Membandingkan tingkat kecocokan butir dari pemodelan GRM, PCM, dan GPCM.

Kecocokan model dapat digunakan sebagai penentu hasil analisis dari model IRT. Dengan semakin banyak butir soal yang cocok dapat dikatakan hasil analisis dari model tersebut akan semakin baik. Untuk mengetahui kecocokan model dapat dilakukan secara statistik (dengan khi-kuadrat) dan secara grafik.

a. Metode Khi-Kuadrat

Metode khi-kuadrat dapat digunakan untuk mencari kecocokan model. Penentuan kecocokan dilakukan dengan melakukan perbandingan antara hasil perhitungan khi-kuadrat dengan khi-kuadrat tabel. Hipotesis pada kecocokannya dapat ditulis sebagai berikut.

$H_0$  : Butir cocok dengan model

$H_1$  : Butir tidak cocok dengan model

Dengan kriteria keputusan  $H_0$  ditolak jika khi-kuadrat hitung lebih besar dibandingkan dengan nilai khi-kuadrat pada tabel. Selain itu, butir juga dapat dikatakan tidak cocok jika p-value kurang dari 0,05. Model yang mempunyai banyaknya butir cocok paling banyak dipilih sebagai model yang paling cocok dan paling baik.

Metode khi-kuadrat yang sering digunakan yaitu dengan Metode *Yen's QI* dengan rumus sebagai berikut (Yen, 1981).

$$Q_{1i} = \sum_{k=1}^{10} \frac{N_k(O_{ik} - E_{ik})^2}{E_{ik}} \tag{6}$$

Dengan derajat bebas  $10 \times (K - 1) - m$  dengan K merupakan banyaknya respons peserta tes dan m merupakan banyaknya parameter yang diestimasi. Dimana:

$Q_{1i}$  : nilai statistik dari butir ke-k.

$O_{ik}$  : proporsi dari yang diamati atau dari peserta yang menjawab benar pada butir-i yang berada di kelompok-k.

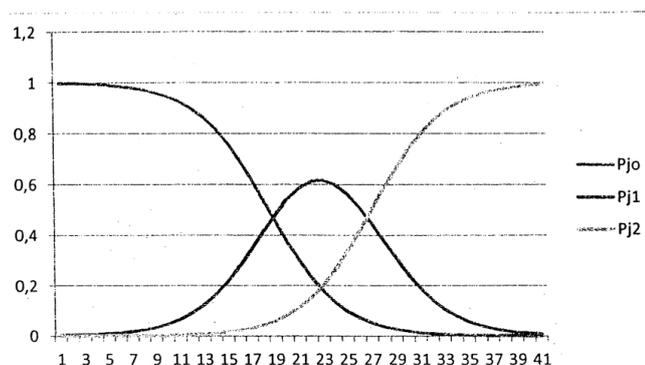
$E_{ik}$  : proporsi yang diharapkan dari peserta yang menjawab benar pada butir-i yang berada di kelompok-k.

$N_k$  : banyaknya peserta di kelompok-k

$Q_{1i}$  : merupakan nilai statistik dari butir ke-k.

b. Secara grafis

Secara grafis kecocokan model dapat ditentungan dengan melihat grafis ICC (Retnawati, 2014). ICC pada pemodelan dikotomi atau dalam politomi dengan CRF. CRF merupakan kependekan dari *Categorical Response Function*. CRF menyatakan hubungan peluang menjawab benar dengan kemampuan peserta tes yang diukur oleh butir soal. Kecocokan model dilihat dengan memanfaatkan nilai observasi dan nilai harapan yang kemudian digambarkan pada kurva. Contoh grafik CRF pada Gambar 3.



Gambar 3. Grafik CRF 3 kategori

Gambar 3 merupakan contoh grafik CRF yang diambil dari Heri Retnawati (2014). Grafik pada Gambar 3 merupakan grafik CRF 3 kategori.

- Menentukan model yang paling fit dengan melihat nilai terkecil dari hasil AIC dan BIC. Selain dilihat banyaknya butir yang cocok, perlu dilihat kriteria dari kecocokan model tersebut. Untuk melihat kriteria model yang cocok dapat dengan melihat nilai AIC dan BIC. AIC merupakan kependekan dari *Akaike Information Criterion*. Persamaan AIC dapat dituliskan sebagai Persamaan (7).

$$AIC = 2k - \log(Likelihood) \tag{7}$$

Dengan k merupakan banyaknya parameter yang akan dilakukan estimasi dan *Likelihood* merupakan nilai *Maximum Likelihood* dari model. Sedangkan BIC merupakan kependekan dari *Bayesian Information Criterion*. Persamaan BIC dapat dituliskan sebagai Persamaan (8).

$$BIC = 2k * \log(n) - \log(Likelihood) \tag{8}$$

Dengan k merupakan banyaknya parameter yang akan dilakukan estimasi, *Likelihood* merupakan nilai *Maximum Likelihood* dari model, dan n merupakan banyaknya pengamatan yang dilakukan. Dari hasil nilai AIC dan BIC tersebut dilihat nilai yang palng kecil sebagai model paling baik.

## HASIL DAN PEMBAHASAN

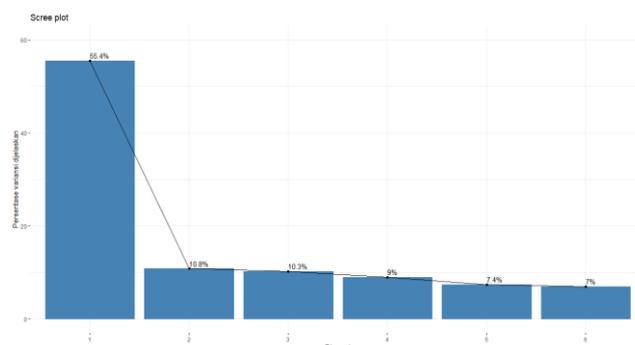
### Hasil

Sebelum melakukan analisis, terlebih dahulu melakukan uji asumsi unidimensi dan independensi lokal pada data. Setelah itu dilakukan estimasi dengan menggunakan model GRM, PCM, dan GPCM terhadap data tersebut. Hasil dari analisis estimasi tersebut selanjutnya dilakukan perbandingan butir yang cocok dengan melihat khi-kuadrat lalu dipilih butir yang mempunyai kecocokan paling banyak. Selain itu juga dilakukan perbandinga nilai AIC dan nilai BIC dimana model terbaik yang mempunuai 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.

#### 1. Uji asumsi

Uji asumsi yang dilakukan pengujian yaitu uji asumsi unidimensi dan independensi lokal. Sedangkan uji asumsi invariansi dapat dilakukan setelah diperoleh model yang paling cocok.

##### a. Unidimensi



Gambar 4. Scree Plot Uji Unidimensi

Dari Gambar 4 dapat dilihat jika nilai eigen pertama menuju nilai eigen kedua terdapat penurunan yang curam. Selain itu, pada nilai eigen kedua terdapat patahan dan pada nilai eigen seterusnya landai sehingga hal ini menandakan jika pada nilai eigen

pertama dominan terhadap faktor lainnya. Hal ini dapat dikatakan jika uji asumsi unidimensi terpenuhi.

b. Independensi lokal

Uji independensi lokal dilakukan dengan melihat korelasi antar butir pada data yang digunakan. Untuk korelasi antar butir dilakukan dengan Aplikasi Rstudio. Hasil korelasi antar butir dapat dilihat pada Tabel 1.

**Tabel 1.** Korelasi Antar Butir

Butir	A	B	C	D	E	F
A	1,00	0,57	0,45	0,55	0,47	0,43
B	0,57	1,00	0,48	0,55	0,52	0,44
C	0,45	0,48	1,00	0,40	0,39	0,39
D	0,55	0,55	0,40	1,00	0,48	0,43
E	0,47	0,52	0,39	0,48	1,00	0,38
F	0,43	0,44	0,39	0,43	0,38	1,00

Dari hasil pada Tabel 1 dapat dilihat jika nilai korelasi antar butir tidak melebihi 0,7. ini memenuhi syarat independensi lokal dengan korelasi kurang dari 0,7. Dari hasil tersebut dapat disimpulkan jika uji asumsi independensi lokal terpenuhi. Selain itu, dengan terbuktinya uji unidimensi juga menandakan secara otomatis uji independensi lokal terpenuhi.

2. Kecocokan model

Untuk mengetahui model yang cocok untuk digunakan, diperlukan uji kecocokan guna mengetahui model yang paling cocok. Dalam penelitian ini dilakukan pemodelan terhadap data yang ada dengan metode GRM, PCM, dan GPCM.

a. GRM

Uji kecocokan model dengan GRM dilakukan dengan menggunakan khi-kuadrat *Yen's QI*. Sebelum melakukan uji khi-kuadrat terlebih dahulu dicari nilai tingkat kesulitan dan nilai diskriminannya. Tingkat kesulitan dilambangkan dengan b dan nilai diskriminannya dilambangkan dengan a. Nilai tingkat kesulitan dan nilai diskriminannya dengan hasil pada Tabel 2.

**Tabel 2.** Estimasi Tingkat Kesulitan(b) dan Diskriminannya(a)

Butir	a	b1	b2	b3	b4	b5	b6	b7	b8	b9
1	2,01	-2,76	-1,77	-0,96	-0,23	0,45	1,36	2,23	3,34	
2	2,25	-2,48	-1,46	-0,74	-0,08	0,55	1,26	1,90	3,01	
3	1,40	-4,35	-2,10	-1,02	-0,06	0,92	1,84	2,84	4,12	
4	1,87	-2,95	-1,90	-1,22	-0,52	0,04	0,65	1,40	2,44	3,95
5	1,55	-2,24	-1,20	-0,35	0,34	1,19	2,09	3,10	4,46	
6	1,37	-4,42	-3,10	-2,05	-1,05	0,06	0,89	1,97	3,10	4,61

Selanjutnya ditentukan nilai kemampuan setiap kelompok dengan nilai kemampuan peserta dilakukan pengelompokan data menjadi 10 kelompok yang sama banyak dan telah dilakukan pengurutan. Dalam penelitian ini setiap kelompok terdapat 85 peserta tes. Dari kemampuan setiap peserta pada satu kelompok dihitung rata-rata kemampuan pada kelompok tersebut. Hasil perhitungan rerata kemampuan setiap kelompok yang disimbolkan dengan  $\theta$  dapat dilihat pada Tabel 3.

**Tabel 3.** Rerata Kemampuan Setiap Kelompok ( $\theta$ )

kelompok	kemampuan setiap kelompok ( $\theta$ )	kelompok	kemampuan setiap kelompok ( $\theta$ )
1	-1,27603	6	-0,04586
2	-0,89605	7	0,17626
3	-0,67429	8	0,49418
4	-0,4749	9	1,00309
5	-0,25654	10	1,91967

Dari Tabel 3 dapat dilihat jika nilai kemampuan setiap kelompok ( $\theta$ ) terurut dari kelompok satu paling kecil sebesar -1,27603 sampai kelompok 10 paling besar dengan nilai 1,91967. Selanjutnya dilakukan penentuan nilai observasi dan prediksi yang nantinya digunakan untuk penentuan nilai khi-kuadrat hitung. Nilai observasi diperoleh dari nilai pada kategori tersebut dibagi total probabilitas. Hasil perhitungan nilai observasi dan prediksi pada butir 1 dapat dilihat pada Tabel 4.

**Tabel 4.** Nilai Proporsi Observasi dan Prediksi Peserta Butir 1 dengan Model GRM

Kelompok	Kategori 0		Kategori 1		...	Kategori 7		Kategori 8	
	$O_{10}$	$E_{10}$	$O_{11}$	$E_{11}$		$O_{17}$	$E_{17}$	$O_{18}$	$E_{18}$
1	0,1059	0,0062	0,2588	0,1498	...	0,0000	0,0000	0,0000	0,0000
2	0,0118	0,0017	0,1294	0,0463	...	0,0000	0,0000	0,0000	0,0000
3	0,0118	0,0008	0,1176	0,0223	...	0,0000	0,0000	0,0000	0,0000
4	0,0000	0,0004	0,0824	0,0114	...	0,0000	0,0001	0,0000	0,0000
5	0,0118	0,0002	0,0706	0,0054	...	0,0000	0,0002	0,0000	0,0000
6	0,0000	0,0001	0,0118	0,0027	...	0,0000	0,0004	0,0000	0,0000
7	0,0000	0,0000	0,0000	0,0012	...	0,0118	0,0009	0,0000	0,0000
8	0,0000	0,0000	0,0000	0,0004	...	0,0588	0,0026	0,0118	0,0001
9	0,0000	0,0000	0,0000	0,0001	...	0,0588	0,0147	0,0118	0,0003
10	0,0000	0,0000	0,0000	0,0000	...	0,3529	0,2524	0,0824	0,0078

Dari hasil nilai observasi dan prediksi butir 1 pada Tabel 4 dapat digunakan untuk melakukan uji khi-kuadrat *Yen's QI*. Perhitungan di atas dilakukan juga pada butir lainnya sehingga diperoleh hasil seperti pada Tabel 5.

**Tabel 5.** Kecocokan dengan Model GRM

Butir	$X_2$	p-value	Kecocokan
1	79.452	0.230	Cocok
2	95.823	0.027	Tidak Cocok
3	65.066	0.676	Cocok
4	89.009	0.230	Cocok
5	59.137	0.841	Cocok
6	95.211	0.118	Cocok

Pada Tabel 5 dapat dilihat jika terdapat 5 butir yang cocok yaitu pada butir 1, butir 3, butir 4, butir 5 dan butir 6. Sedangkan untuk butir yang tidak cocok terdapat 1 butir yaitu butir 2.

b. PCM

Pada dasarnya proses yang dilakukan untuk melakukan uji kecocokan dengan PCM sama dengan yang dilakukan pada Model GRM tetapi hanya berbeda pada probabilitasnya saja. Sama dengan GRM, nilai tingkat kesulitan dan nilai diskriminannya dapat dilihat pada Tabel 12.

**Tabel 6.** Estimasi Tingkat Kesulitan(b) dan Diskriminannya(a)

Butir	a	b1	b2	b3	b4	b5	b6	b7	b8	b9
1	1	-2,38	-1,46	-0,69	-0,08	0,21	1,18	1,58	2,74	
2	1	-2,48	-1,15	-0,52	0,04	0,44	1,12	1,08	2,65	
3	1	-3,67	-1,31	-0,69	-0,05	0,72	1,22	1,75	2,65	
4	1	-2,40	-1,28	-1,00	-0,13	0,00	0,35	0,85	1,89	3,62
5	1	-1,54	-0,80	-0,04	0,10	0,86	1,41	2,05	3,21	
6	1	-2,55	-1,91	-1,41	-0,74	0,02	0,56	1,36	1,96	3,12

Dari Tabel 6 dapat dilihat jika nilai diskriminan pada setiap butir sama yaitu bernilai 1 hal ini sesuai dengan model PCM yang diasumsikan jika setiap butir mempunyai nilai yang sama.

Selanjutnya sama dengan melakukan perhitungan rerata kemampuan setiap kelompok ( $\theta$ ) dengan cara yang sama seperti pada model GRM. Hasil perhitungan rerata kemampuan setiap kelompok ( $\theta$ ) dapat dilihat pada Tabel 7.

**Tabel 7.** Rerata Kemampuan Setiap kelompok ( $\theta$ )

Kelompok	Kemampuan setiap kelompok ( $\theta$ )	Kelompok	Kemampuan setiap kelompok ( $\theta$ )
1	-0,94261	6	-0,02903
2	-0,63499	7	0,14061
3	-0,47856	8	0,36906
4	-0,32065	9	0,70860
5	-0,18062	10	1,36677

Dari Tabel 7 dapat dilihat jika nilai kemampuan setiap kelompok ( $\theta$ ) terurut dari kelompok satu paling kecil sebesar -0,94261 sampai kelompok 10 paling besar dengan nilai 1,36677.

Selanjutnya sama dengan model GRM dengan mencari nilai observasi dan prediksi. Untuk nilai observasi sama dengan model GRM tetapi untuk nilai prediksi berbeda dengan model PCM. Perhitungan prediksi proporsi pada model PCM dengan rumus pada persamaan 3. Contoh hasil perhitungan nilai observasi dan prediksi pada butir 1 dapat dilihat pada Tabel 8.

**Tabel 8.** Nilai Proporsi Observasi dan Prediksi Peserta Butir 1 dengan Model PCM

Kelompok	Kategori 0		Kategori 1		...	Kategori 7		Kategori 8	
	$O_{10}$	$E_{10}$	$O_{11}$	$E_{11}$		$O_{17}$	$E_{17}$	$O_{18}$	$E_{18}$
1	0,1059	0,0481	0,2588	0,2019	...	0,0000	0,0003	0,0000	0,0000
2	0,0118	0,0221	0,1294	0,1261	...	0,0000	0,0013	0,0000	0,0000
3	0,0118	0,0141	0,1176	0,0940	...	0,0000	0,0025	0,0000	0,0001
4	0,0000	0,0086	0,0824	0,0671	...	0,0000	0,0046	0,0000	0,0002
5	0,0118	0,0054	0,0706	0,0481	...	0,0000	0,0077	0,0000	0,0004
6	0,0000	0,0031	0,0118	0,0324	...	0,0000	0,0128	0,0000	0,0008
7	0,0000	0,0016	0,0000	0,0198	...	0,0118	0,0217	0,0000	0,0016
8	0,0000	0,0006	0,0000	0,0095	...	0,0588	0,0410	0,0118	0,0038
9	0,0000	0,0001	0,0000	0,0027	...	0,0588	0,0902	0,0118	0,0118
10	0,0000	0,0000	0,0000	0,0001	...	0,3529	0,2537	0,0824	0,0643

Tabel 8 merupakan nilai observasi dan prediksi dari setiap kategori dalam setiap kelompok pada butir 1.

Hasil nilai observasi dan prediksi butir 1 pada Tabel 8 juga digunakan untuk melakukan uji khi-kuadrat *Yen's QI*. Perhitungan khi-kuadrat dilakukan juga pada butir lainnya sehingga diperoleh hasil seperti pada Tabel 9.

**Tabel 1.** Kecocokan dengan Model PCM

Butir	$X_2$	p-value	Kecocokan
1	108.077	0.004	Tidak Cocok
2	121.958	0.000	Tidak Cocok
3	85.402	0.134	Cocok
4	93.455	0.163	Cocok
5	66.161	0.672	Cocok
6	88.123	0.276	Cocok

Pada Tabel 9 dapat dilihat jika terdapat 4 butir yang cocok yaitu pada butir 3, butir 4, butir 5 dan butir 6. Sedangkan untuk butir yang tidak cocok terdapat 2 butir yaitu pada butir 1, dan butir 2.

c. GPCM

GPCM merupakan pengembangan dari model PCM. Dalam melakukan uji kecocokan dengan model GPCM sama dengan yang dilakukan pada model PCM tetapi hanya berbeda pada probabilitasnya dengan terdapat D yang merupakan faktor skala sebesar 1,7 dan diskriminan (a) pada model. Nilai tingkat kesulitan dan nilai diskriminannya dapat dilihat pada Tabel 10.

**Tabel 2.** Estimasi Tingkat Kesulitan(b) dan Diskriminannya(a)

Butir	a	b1	b2	b3	b4	b5	b6	b7	b8	b9
1	1	-2,81	-1,76	-0,88	-0,15	0,28	1,43	2,04	3,39	
2	1	-2,85	-1,42	-0,68	-0,01	0,52	1,35	1,51	3,24	
3	1	-5,93	-1,98	-1,02	-0,05	1,13	1,80	2,51	3,87	
4	1	-3,14	-1,69	-1,32	-0,21	-0,02	0,46	1,15	2,51	4,72
5	1	-2,23	-1,14	-0,02	0,13	1,23	1,99	2,88	4,57	
6	1	-3,92	-2,92	-2,17	-1,13	0,06	0,86	2,05	2,87	4,66

Dari Tabel 10 dapat dilihat jika nilai diskriminan pada setiap butir sama dengan PCM yaitu bernilai 1 dan nilai estimasi tingkat kesulitan semakin ke kanan semakin besar.

Perolehan nilai rerata kemampuan setiap kelompok ( $\theta$ ) sama dengan model-model sebelumnya. Hasil perhitungan rerata kemampuan setiap kelompok ( $\theta$ ) dapat dilihat pada Tabel 11.

**Tabel 3.** Rerata Kemampuan Setiap kelompok ( $\theta$ )

kelompok	kemampuan setiap kelompok ( $\theta$ )	kelompok	kemampuan setiap kelompok ( $\theta$ )
1	-1,28665	6	-0,03775
2	-0,88015	7	0,18674
3	-0,65657	8	0,49970
4	-0,46013	9	0,97844
5	-0,25187	10	1,85580

Dari Tabel 11 dapat dilihat jika nilai kemampuan setiap kelompok ( $\theta$ ) terurut dari kelompok 1 paling kecil sebesar -1,28665 sampai kelompok 10 paling besar dengan nilai 1, 85580.

Nilai observasi pada model GPCM sama dengan model-model sebelumnya. Sedangkan nilai prediksi pada model GPCM mempunyai persamaan yang sama dengan model PCM tetapi pada model GPCM ditambah nilai faktor skala (D) sebesar

1,7 dan dikalikan dengan nilai diskriminannya(a). Perhitungan nilai proporsi prediksi pada model GPCM dengan rumus pada persamaan 7. Hasil perhitungan nilai observasi dan prediksi dapat dilihat pada Tabel 12.

**Tabel 4.** Nilai Proporsi Observasi dan Prediksi Peserta Butir 1 dengan Model GPCM

Kelom pok	Kategori 0		Kategori 1		...	Kategori 7		Kategori 8	
	$O_{10}$	$E_{10}$	$O_{11}$	$E_{11}$		$O_{17}$	$E_{17}$	$O_{18}$	$E_{18}$
1	0,1059	0,0165	0,2588	0,2165	...	0,0000	0,000000	0,0000	0,000000
2	0,0118	0,0033	0,1294	0,0871	...	0,0000	0,000002	0,0000	0,000000
3	0,0118	0,0012	0,1176	0,0453	...	0,0000	0,000011	0,0000	0,000000
4	0,0000	0,0004	0,0824	0,0232	...	0,0000	0,000043	0,0000	0,000000
5	0,0118	0,0001	0,0706	0,0103	...	0,0000	0,000159	0,0000	0,000000
6	0,0000	0,0000	0,0118	0,0039	...	0,0000	0,000541	0,0000	0,000002
7	0,0000	0,0000	0,0000	0,0013	...	0,0118	0,001725	0,0000	0,000008
8	0,0000	0,0000	0,0000	0,0002	...	0,0588	0,007079	0,0118	0,000052
9	0,0000	0,0000	0,0000	0,0000	...	0,0588	0,040857	0,0118	0,000682
10	0,0000	0,0000	0,0000	0,0000	...	0,3529	0,318404	0,0824	0,023549

Tabel 12 merupakan nilai observasi dan prediksi dari setiap kategori dalam setiap kelompok pada butir 1.

Hasil nilai observasi dan prediksi pada butir 1 pada Tabel 12 digunakan untuk melakukan uji khi-kuadrat *Yen's QI*. Perhitungan khi-kuadrat dilakukan juga pada butir lainnya sehingga diperoleh hasil seperti pada Tabel 13.

**Tabel 5.** Kecocokan dengan Model GPCM

Butir	X2	p-value	Kecocokan
1	82.534	0.165	Cocok
2	96.199	0.025	Tidak Cocok
3	73.907	0.383	Cocok
4	79.232	0.503	Cocok
5	60.723	0.803	Cocok
6	99.019	0.073	Cocok

Pada Tabel 13 dapat dilihat jika terdapat 5 butir yang cocok yaitu pada butir 1, butir 3, butir 4, butir 5, dan butir 6. Sedangkan untuk butir yang tidak cocok terdapat 1 butir yaitu pada butir 2.

### 3. Model paling cocok

Pada penelitian ini terdapat 3 model politomi yang digunakan yaitu GRM, PCM, dan GPCM. Dimana dari ketiga model tersebut telah dilakukan uji kecocokan model dengan melihat nilai khi-kuadrat *Yen's QI*. Perbandingan kecocokan model dari ketiga model tersebut dapat dilihat pada Tabel 14.

**Tabel 14.** Perbandingan Butir yang Cocok dari Model GRM, PCM, dan GPCM

Butir	GRM	PCM	GPCM
1	Cocok	Tidak Cocok	Cocok
2	Tidak Cocok	Tidak Cocok	Tidak Cocok
3	Cocok	Cocok	Cocok
4	Cocok	Cocok	Cocok
5	Cocok	Cocok	Cocok
6	Cocok	Cocok	Cocok

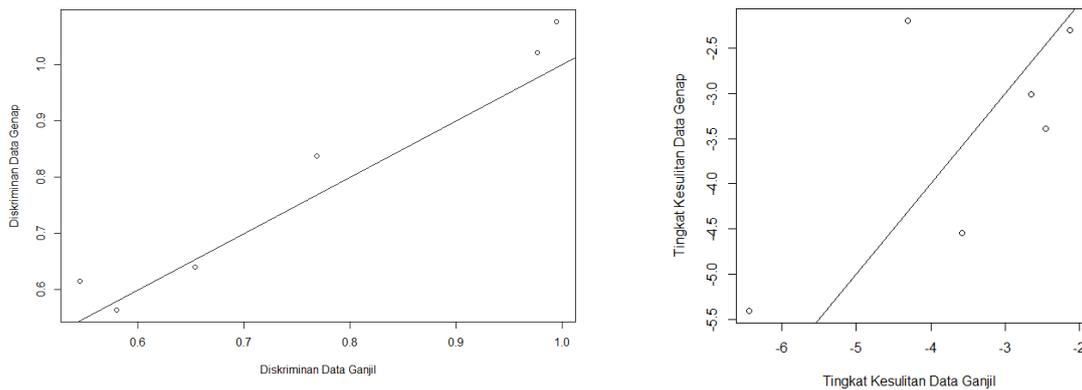
AIC dan BIC digunakan karena menurut Linden (2016) AIC dan BIC digunakan untuk menemukan satu model yang terbaik pada sebagian besar penggunaan model IRT. Hasil perbandingan melihat nilai AIC dan nilai BIC model GRM, PCM, dan GPCM dapat dilihat pada Tabel 15.

**Tabel 15.** Perbandingan Kriteria dari Kecocokan Model Model GRM, PCM, Dan GPCM

Model	Butir yang cocok	AIC	BIC
GRM	5	18312,77	18578,57
PCM	2	18329,39	18571,45
GPCM	5	<b>18290,85</b>	<b>18556,65</b>

Dari Tabel 14 dan 15 dapat dilihat jika model GRM dan model GPCM mempunyai banyak butir yang cocok sama dan paling banyak yaitu 5. Akan tetapi, nilai AIC paling kecil terdapat pada model GPCM sebesar 18290,85. Sedangkan nilai BIC paling kecil juga terdapat pada model GPCM dengan nilai sebesar 18556,65. Sehingga model yang paling baik dan paling cocok yaitu model GPCM.

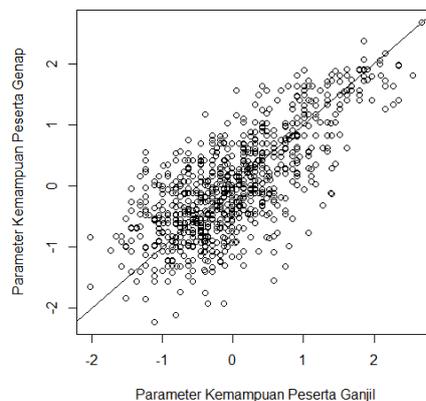
4. Uji asumsi invariansi parameter



**Gambar 5.** Plot Invariansi Parameter Diskriminan dan Tingkat Kesulitan

Dari Gambar 5 dapat dilihat jika pada invariansi parameter diskriminan titik-titiknya berada pada sekitar garis atau bisa dikatakan mengikuti garis. pada invariansi parameter tingkat kesulitan titik-titiknya mayoritas berada pada garis meskipun terdapat titik yang menjauh dari garis. Dari hal ini dapat dikatakan jika uji invariansi parameter pada diskriminan dan tingkat kesulitan terpenuhi.

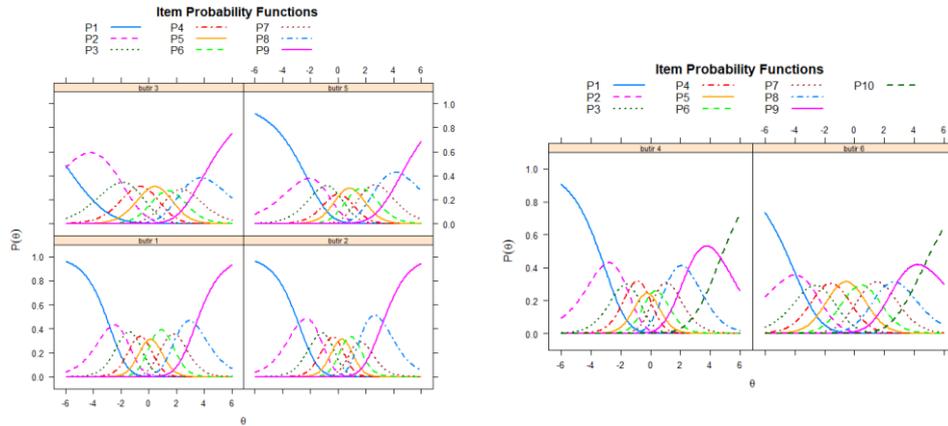
Selain kedua invariansi di atas, juga dilihat invariansi parameter kemampuan ( $\theta$ ) dengan membagi butir soal menjadi butir nomor ganjil dan butir nomor genap dengan hasil plot invariansi parameter kemampuan dapat dilihat pada Gambar 6.



**Gambar 6.** Invariansi Parameter Kemampuan

Dari Gambar 6 dapat dilihat jika titik-titik berada di sekitar garis atau bisa dikatakan mengikuti gatis. Hal ini menandakan bahwa invariansi parameter kemampuan ( $\theta$ ) terpenuhi.

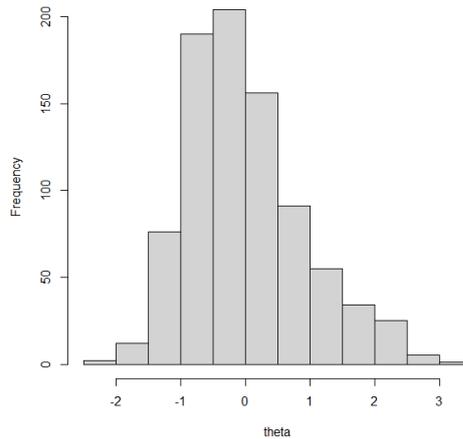
5. Plot CRF dengan model paling baik



**Gambar 7.** Grafik CRF dengan Model Terbaik

Pada Gambar 7 menunjukkan jika butir dapat digambarkan oleh model. Semakin ke kanan menunjukkan peluang peserta mampu dalam kategori-kategori semakin tinggi kemampuannya. Grafik butir 1 pada titik potong antara garis P1 dan P2 menunjukkan jika tingkat kesulitan menjawab kategori 1 sebesar -2,8. Pada titik potong antara garis P2 dan P3 menunjukkan jika tingkat kesulitan menjawab kategori 2 sebesar -1,8, begitu seterusnya sampai semua kategori dan semua butir. Semakin nilai kemampuan mendekati puncak pada grafik maka probabilitas peserta memperoleh skor pada kategori tersebut semakin tinggi.

6. Estimasi kemampuan dengan model paling baik



**Gambar 8.** Histogram Estimasi Kemampuan dengan Model Paling Baik

Dari Gambar 8 dapat dilihat jika kemampuan peserta dalam mengerjakan tes pengukur kemampuan berbahasa Inggris dengan model GPCM sebarannya berada mendekati nilai kemampuan rerata yaitu di sekitar nilai nol. Dari hasil tersebut dapat dikatakan jika kemampuan peserta dalam mengerjakan tes pengukur kemampuan berbahasa Inggris dengan model GPCM tersebar secara normal.

**Pembahasan**

Penelitian ini mempunyai tujuan untuk membandingkan kecocokan model GRM, PCM, dan GPCM pada data tes *reading* dalam uji kemampuan berbahasa Inggris yang diskor politomi. Perbandingan dilakukan dengan uji khi-kuadrat metode *Yen's QI* dan dengan melihat nilai AIC dan BIC. Model yang mempunyai butir cocok paling banyak dan nilai AIC dan BIC paling rendah dipilih menjadi model paling baik dan paling cocok untuk digunakan.

Pada uji khi-kuadrat *Yen's QI* dengan membagi data tes pengukur kemampuan berbahasa Inggris menjadi 10 kelompok secara terurut. Hasil khi-kuadrat diperoleh hasil perbandingan dengan model dengan jumlah butir yang cocok paling banyak terdapat pada model GRM dan model GPCM yaitu mempunyai banyak butir yang cocok masing-masing sama yaitu berjumlah 5. Dari hasil tersebut, dilihat nilai AIC dan BIC pada model GRM dan model GPCM. Pada Tabel 21, nilai AIC dan BIC yang paling kecil sama-sama terdapat pada model GPCM. Sehingga model GPCM dipilih sebagai model yang paling baik dengan 5 butir yang cocok, nilai AIC sebesar 18290,85, dan nilai BIC sebesar 18556,65.

Hal ini selaras dengan penelitian yang dilakukan oleh Saiful Ridlo (2012) yang melakukan perbandingan kecocokan pada model GRM dan GPCM dalam pemodelan skor pada Tes Pengetahuan Praktikum Biologi (TPPB) dengan hasil jika model GPCM lebih cocok untuk digunakan. Akan tetapi, hasil penelitian ini berbeda dengan yang dilakukan oleh Auné et al. (2020) pada data skala tingkat kesepian dengan hasil jika model GRM lebih baik daripada model lain yang dibandingkan. Hal ini menunjukkan jika pada data hasil tes, untuk metode penskoran yang lebih cocok dengan menggunakan model GPCM. Hasil penelitian ini berbeda dikarenakan karakteristik dari data yang digunakan berupa data skala angket dan bukan pengukur kemampuan.

Pada model terbaik juga dilihat grafik CRF yang menunjukkan butir dapat digambarkan oleh model. Semakin ke kanan menunjukkan peluang peserta mampu dalam kategori-kategori semakin tinggi kemampuannya. Pada grafik, titik potong antara garis probabilitas kategori menunjukkan tingkat kesulitan menjawab pada kategori. Grafik CRF memudahkan dalam menunjukkan kemungkinan peserta dalam memilih kategori tertentu. Hal serupa juga digunakan oleh Dewanti et al. (2021) dalam penelitiannya yang berjudul *The Application of Item Response Theory in Analysis of Characteristics of Mathematical Literacy Test Items*. Mereka melakukan analisis kemampuan dalam literasi matematika dari 258 siswa menggunakan teori respons butir dengan pendekatan GPCM. Hasil dari penelitian tersebut menunjukkan jika model GPCM cocok untuk menganalisis instrumen tes literasi matematika dengan 4 level literasi matematika.

Selain itu juga dilihat estimasi kemampuan pada model yang paling baik dengan hasil jika kemampuan peserta dalam mengerjakan tes pengukur kemampuan berbahasa Inggris dengan model GPCM tersebar secara normal. Penggunaan estimasi kemampuan juga diterapkan oleh Saepuzaman et al. (2021) dalam penelitiannya yang membandingkan kemampuan siswa dengan penskoran dikotomis model 2PL serta pada politomis model GPCM teori respons butir. Penelitian ini menggunakan data 1175 siswa SMA kelas XI yang tersebar di provinsi Jawa Barat dan Banten. Hasil grafik estimasi kemampuan pada penelitian ini menunjukkan pada dikotomis model 2PL sebaran kemampuan peserta tes terbanyak di bawah rata-rata kemampuan. Pada penskoran politomis model GPCM tersebar normal karena penskoran rata-rata kemampuan peserta tes berada disekitar nol.

## SIMPULAN

Berdasarkan hasil analisis kecocokan model teori respons butir pada data jawaban peserta tes terhadap soal *reading* yang mengukur kemampuan mahasiswa dalam berbahasa Inggris dapat diperoleh kesimpulan sebagai berikut:

1. Kecocokan model GRM pada data tes *reading* bahasa Inggris yang diskor politomi diperoleh hasil sebanyak 5 butir yang cocok yaitu pada butir 1, butir 3, butir 4, butir 4 dan

- butir 6. Selain itu terdapat butir yang tidak cocok yaitu pada butir 2. Nilai AIC dan BIC pada model GRM yaitu 18312,77 dan 18578,57.
2. Kecocokan model PCM pada data tes *reading* bahasa Inggris yang diskor politomi diperoleh hasil terdapat 4 butir yang cocok yaitu pada butir 3, butir 4, butir 5 dan butir 6. Selain itu terdapat 2 butir yang tidak cocok pada butir 1 dan butir 2. Nilai AIC dan BIC pada model PCM yaitu 18329,39 dan 18571,45.
  3. Kecocokan model GPCM pada data tes *reading* bahasa Inggris yang diskor politomi diperoleh hasil sebanyak 5 butir yang cocok yaitu pada butir 1, butir 3, butir 4, butir 5, dan butir 6. Sedangkan butir yang tidak cocok pada butir 2. Nilai AIC dan BIC pada model GPCM yaitu 18290,85 dan 18556,65.
  4. Berdasarkan perbandingan kecocokan model GRM, PCM, dan GPCM pada data tes *reading* bahasa Inggris yang diskor politomi yang telah diutarakan pada poin 1 sampai 3, diperoleh model yang paling baik dan paling cocok untuk digunakan adalah model GPCM. Model GPCM dipilih karena mempunyai butir cocok yang paling banyak dan mempunyai nilai AIC dan BIC yang terkecil.

## DAFTAR PUSTAKA

- Adiyo. (2014). Pengaruh Jumlah Kategori, Rentang Threshold dan Metode Estimasi Terhadap Akurasi Skor Tes Pada Beberapa Model Politomi. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 3, 691–729.
- Auné, S. E., Juan, F., Abal, P., & Attorresi, H. F. (2020). *A Psychometric Analysis from The Item Response Theory: Step-By-Step Modelling of A Loneliness Scale*. *Ciencias Psicológicas*, 14(June), 1–15.
- DeMars, C. (2010). *Item Response Theory: Understanding Statistic Measurement*. New York: Oxford University Press, Inc.
- Desjardins, C. D., & Bulut, O. (2018). *Handbook of Educational Measurement and Psychometrics Using R*. New York: CRC Press.
- Dewanti, S. S., Hadi, S., Nu'man, M., & Ibrahim. (2021). *The Application of Item Response Theory in Analysis of Characteristics of Mathematical Literacy Test Items*. *Ilkogretim Online - Elementary Education Online*, 20(1), 1226–1237.
- Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1991). *Fundamentals of Item Response Theory*. California: SAGE Publications.
- Kadir Abdul. (2015). Menyusun dan Menganalisis Tes Hasil Belajar. *Jurnal Al-Ta'dib*, 8(2), 70–81.
- Linden, W. J. van der. (2016). *Handbook of Item Response Theory: Statistical Tools*. Florida: CRC Press.
- Retnawati, H. (2014). Teori Respons Butir dan Penerapannya: Untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana. Yogyakarta: Nuha Medika.
- Saepuzaman, D., Istiyono, E., -, H., Retnawati, H., & -, Y. (2021). Analisis Estimasi Kemampuan Siswa Dengan Pendekatan Item Response Theory Penskoran Dikotomus Dan Politomus. *Karst : JURNAL PENDIDIKAN FISIKA DAN TERAPANNYA*, 4(1), 8–13.
- Safaruddin, Anisa, & AF, M. S. (2012). Partial Credit Model ( PCM ) dalam Penskoran Politomi pada Teori Respon Butir. *Jurnal Matematika, Statistika, & Komputasi*, 9(1), 39–48.
- Sudaryono. (2011). Implementasi Teori Responsi Butir (Item Response Theory) Pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan Dan Kebudayaan*, 17(6), 719–732.
- Yen, W. M. (1981). Using Simulation Results to Choose a Latent Trait Model. *Applied Psychological Measurement*, 5(2), 245–262.
- Zainal, M. A., Effendi, M., & Matore, E. M. (2020). a Rasch Model Analysis on Teachers'

Innovative Behaviour Psychometric Items. *Journal of Mechanics of Continua and Mathematical Sciences*, 15(8), 265–276.