

Aplikasi IRT dalam Analisis Aitem Tes Kognitif

Firmanto Adi Nurcahyo¹

Fakultas Psikologi Universitas Pelita Harapan Surabaya

Abstract

Item Response Theory (IRT) was developed to overcome the problems in Classical Test Theory (CTT). Item analysis using IRT is based on the parameters used which are item difficulty, item discrimination, and pseudo-chance level. This paper aims to provide an overview of how IRT is used to perform item analysis on cognitive tests. Simulation data of seven items with 900 subjects were used. IRT analysis was performed with R Program based on one, two, and three-parameter logistic model. The results were then discussed.

Keywords: item analysis, item difficulty, Item Response Theory

Pengantar

Classical Test Theory (CTT) atau teori tes klasik telah banyak digunakan dalam proses analisis aitem. Popularitas ini bisa jadi dikarenakan kelebihan yang dimiliki CTT. Kelebihan tersebut misalnya taraf kesukaran dan daya diskriminasi aitem dalam teori tes klasik dapat dihitung secara manual. Penghitungan yang dapat dikerjakan secara manual dikarenakan analisis dengan CTT didasarkan pada data dengan jumlah yang tidak terlalu banyak.

Selain kelebihan, CTT juga tidak lepas dari kelemahan. Kelemahan tersebut misalnya taraf kesukaran dan daya diskriminasi aitem yang diperoleh bergantung pada sampel (Hambleton & Swaminathan, 1985). Ketergantungan terhadap sampel menyebabkan karakteristik aitem yang dianalisis dengan CTT dapat berubah sesuai konteks dari responden. Artinya, suatu aitem bisa memiliki taraf kesukaran rendah karena aitem tersebut dikerjakan oleh kelompok responden dengan kemampuan tinggi. Namun demikian, taraf kesukaran aitem

tersebut bisa menjadi tinggi ketika dikerjakan oleh kelompok responden dengan kemampuan rendah.

Kelemahan lain dari teori tes klasik adalah teori ini lebih berorientasi pada tes dibandingkan aitem (Hambleton, Swaminathan, & Rogers, 1991). Teori tes klasik tidak memperhatikan bagaimana respons dari responden terhadap aitem. Pada penerapannya, kemampuan responden dilihat berdasarkan skor total dari jumlah jawaban benar dari responden, tanpa membedakan apakah aitem yang dijawab benar oleh responden merupakan aitem yang mudah atau sukar.

Kelemahan-kelemahan dari teori tes klasik memicu lahirnya *Item Response Theory* (IRT) atau teori respons butir. IRT merupakan kerangka umum dari fungsi matematika yang menjelaskan interaksi antara subjek dan butir tes (Sumintono & Widhiarso, 2013). Estimasi terhadap parameter aitem atau abilitas responden pada IRT tidak bergantung pada sampel aitem tertentu atau responden yang dipilih dalam suatu tes.

¹ Korespondensi mengenai artikel ini dapat dilakukan melalui: firmanto.adicahyo@uph.edu

IRT telah banyak mengalami perkembangan. Penerapan IRT pada saat ini tidak hanya dikenakan pada tes yang bersifat unidimensional, namun telah merambah pada tes multidimensional. Azwar dan Ridho (2013) misalnya menggunakan IRT multidimensional untuk mengetahui karakteristik aitem tes potensi. Selain itu, IRT yang pada awalnya dikembangkan untuk melakukan analisa tes kognitif telah juga digunakan pada skala psikologi. Widhiarso (2016) misalnya menggunakan analisis Rasch, yang pada hakekatnya merupakan analisis IRT model satu parameter, untuk mengidentifikasi proporsi responden yang memiliki gaya respons ekstrim saat mengisi skala harga diri.

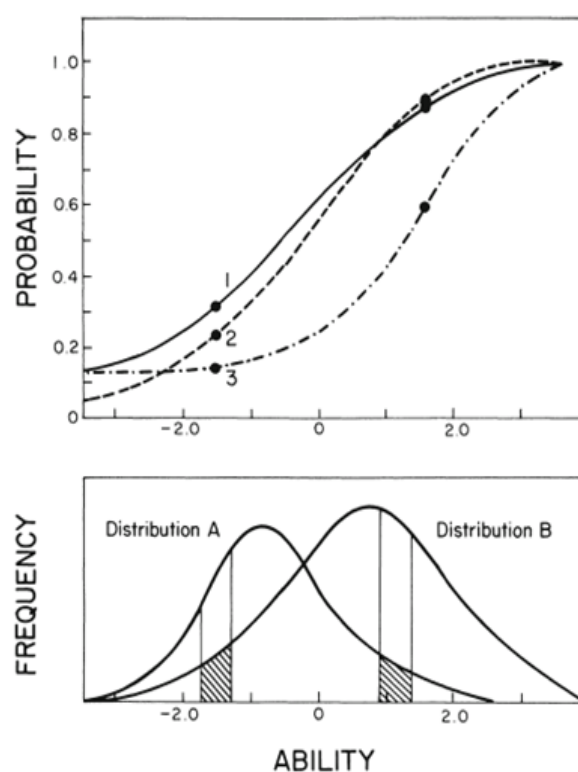
Tulisan ini bertujuan untuk memberikan gambaran analisis aitem berdasarkan IRT model logistik satu, dua, dan tiga parameter yang diterapkan pada tes kognitif. Untuk itu, paparan mengenai IRT secara teoretis akan diberikan sebelumnya. Tulisan ini bermanfaat untuk mengenalkan IRT kepada pembaca, khususnya bagi pembaca yang jarang bersinggungan dengan area psikometri. Selain itu, tulisan ini bermanfaat untuk memberikan gambaran penerapan IRT dalam analisis aitem.

Pembahasan

Item Response Theory (IRT)

IRT dibangun dari pemahaman bahwa probabilitas responden menjawab benar terhadap suatu aitem dapat dideskripsikan sebagai fungsi sederhana dari posisi responden pada suatu trait laten, ditambah dengan satu atau lebih parameter yang menjadi karakteristik aitem (Molenaar, 1995). Hambleton, *et al.* (1991) menyatakan bahwa dasar dari IRT adalah (a) performansi responden terhadap aitem dapat diprediksikan berdasar sejumlah faktor yang disebut trait atau abilitas laten yang

menunjukkan kemampuan atau ciri sifat, (b) hubungan performansi responden terhadap aitem dan trait yang mendasari performansi terhadap aitem digambarkan meningkat secara monotonik, membentuk suatu fungsi yang disebut *Item Characteristic Curve (ICC)*. ICC dari tiga aitem terlihat pada Gambar 1. Dari gambar tersebut dapat diketahui bahwa probabilitas responden dalam menjawab aitem dengan benar bergantung pada abilitas, tanpa membedakan apakah responden berasal dari kelompok A atau B.

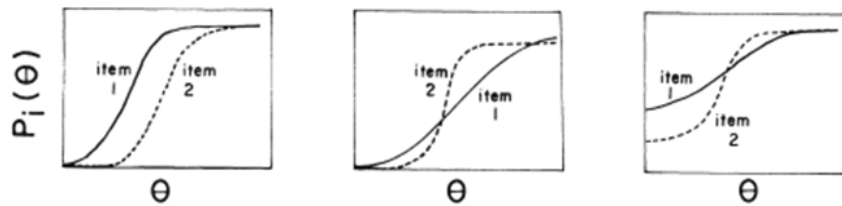


Gambar 1. Probabilitas Menjawab 3 Aitem pada Kelompok A & B (Hambleton & Swaminathan, 1985)

Model Logistik Satu, Dua, dan Tiga Parameter

Model yang populer digunakan dalam IRT adalah model logistik satu, dua, dan tiga parameter (Hambleton *et al.*, 1991). Nama model tersebut disesuaikan dengan jumlah parameter aitem yang dipergunakan. Parameter yang dimaksud adalah taraf kesukaran aitem, daya diskriminasi aitem,

dan tebakan semu. Gambar 2 menunjukkan ICC dua aitem pada model logistik satu, dua, dan tiga parameter.



Gambar 2. ICC pada Model Satu, Dua, dan Tiga Parameter (Hambleton & Swaminathan, 1985)

ICC model logistik satu parameter dibangun dari persamaan:

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}}$$

$P_i(\theta)$ adalah probabilitas responden yang terpilih secara random dengan kemampuan θ menjawab aitem i dengan benar, b_i adalah parameter kesukaran aitem i , dan e adalah nilai 2,718. Parameter b_i merupakan titik pada kontinum abilitas dimana probabilitas respons benar adalah 0,5. Semakin besar nilai parameter b_i , semakin besar pula abilitas yang dibutuhkan bagi responden untuk memperoleh peluang 50% menjawab aitem dengan benar (Hambleton *et al.*, 1991).

ICC untuk model logistik dua parameter yang dikembangkan oleh Birnbaum adalah sebagai berikut:

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

$D = 1.7$ dan a_i merupakan parameter diskriminasi aitem. Parameter ini memberikan informasi sejauh mana aitem mampu membedakan kelompok responden dengan abilitas tinggi dan rendah. Aitem dengan a_i tinggi ditunjukkan dengan kemiringan yang curam. Aitem tersebut lebih mampu membedakan responden pada tingkat abilitas yang berbeda dibandingkan aitem dengan kemiringan yang landai (Hambleton *et al.*, 1991).

Perumusan matematis dari model logistik tiga parameter adalah sebagai berikut:

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

Pada model logistik tiga parameter terdapat tambahan parameter yakni c_i atau *pseudo-chance level* (Hambleton *et al.*, 1991). Parameter ini menunjukkan probabilitas responden dengan kemampuan rendah menjawab aitem dengan benar.

Model logistik satu parameter atau yang terkenal dengan nama model Rasch, merupakan model IRT yang paling sering digunakan. Model ini mengasumsikan bahwa semua aitem mendiskriminasi secara sama serta tidak dapat dijawab dengan benar berdasarkan tebakan (Lord, 1980). Model logistik satu parameter berdasar pada asumsi-asumsi yang restriktif (Hambleton *et al.*, 1991). Kecocokan asumsi bergantung pada data yakni misalnya pada tes yang relatif mudah dan terdiri dari aitem-aitem homogen. Dalam kondisi ukuran sampel yang kecil, estimasi yang dihasilkan model Rasch dimungkinkan lebih akurat jika dibandingkan dengan hasil dari model tiga parameter (Lord, 1980). Model ini juga dinilai memiliki kemudahan dalam pelaksanaannya karena jumlah parameter yang sedikit (Hambleton & Swaminathan, 1985).

Model logistik dua parameter merupakan generalisasi dari model satu parameter yang memungkinkan adanya perbedaan pada daya diskriminasi aitem. Model ini dapat dikenakan pada tes dengan aitem yang direspon secara bebas. Selain itu, model logistik dua parameter juga dapat dikenakan pada tes pilihan ganda, dalam kondisi tes tersebut tidak terlalu sukar bagi individu (Hambleton *et al.*, 1991).

Model logistik tiga parameter cocok dikenakan pada tes yang memandang tebakan sebagai faktor yang berkontribusi penting dalam performansi tes. Kondisi seperti ini dapat terjadi pada tes pilihan ganda. Parameter tebakan semu pada umumnya terdeteksi pada tes dengan aitem-aitem yang memiliki taraf kesukaran yang tinggi. Aitem-aitem yang sukar untuk dijawab memungkinkan individu untuk memilih jawaban dengan cara menebak (Hambleton *et al.*, 1991).

Terdapat hal-hal yang perlu diperhatikan dalam pemilihan model logistik satu, dua atau tiga parameter ketika melakukan analisa data berdasarkan IRT. Wiberg (2004) mengungkapkan bahwa jika diskriminasi aitem ditemukan berbeda-beda pada tiap aitem, maka parameter tersebut perlu dimasukkan dalam model. Hal ini berarti model logistik satu parameter kurang cocok untuk dipakai dalam analisis. Parameter tebakan semu juga sebaiknya dimasukkan pada model jika didapati responden dengan kemampuan rendah masih bisa mendapatkan skor benar pada aitem yang sulit. Pelibatan tebakan semu berarti menekankan penggunaan model logistik tiga parameter.

Jenis tes juga perlu diperhatikan dalam pemilihan model IRT yang dipakai dalam analisis. Hal ini dapat dilukiskan pada tes yang memberikan pengurangan nilai jika responden menjawab salah. Pada kondisi tersebut analisis tiga parameter bisa jadi

kurang tepat diterapkan karena kemungkinan responden untuk menebak cenderung kecil. Sebaliknya, pada tes yang merupakan bentuk *speed test*, analisis tiga parameter dimungkinkan justru lebih tepat dipakai. Hal ini dikarenakan dalam kondisi waktu yang terbatas, kecenderungan responden untuk menebak jawaban cukup tinggi.

Fungsi Informasi

Salah satu hal yang perlu diperhatikan dalam analisis aitem dengan IRT adalah fungsi informasi aitem. Fungsi informasi aitem menunjukkan kontribusi yang diberikan aitem terhadap estimasi kemampuan pada suatu titik dalam kontinum kemampuan. Fungsi informasi tinggi jika taraf kesukaran aitem mendekati kemampuan, daya diskriminasi aitem tinggi, serta peluang menebak mendekati nol (Hambleton *et al.*, 1991).

Gabungan fungsi informasi dari keseluruhan aitem akan membentuk fungsi informasi tes. Hambleton dan Swaminathan (1985) menunjukkan bahwa karakteristik fungsi informasi tes adalah (a) ditetapkan untuk satu set aitem tes pada setiap titik dari kontinum abilitas, (b) jumlah informasi dipengaruhi oleh kualitas dan jumlah aitem tes, (c) kontribusi setiap aitem tidak tergantung pada aitem lainnya, dan (d) jumlah informasi dari satu set aitem tes pada suatu tingkat kemampuan berbanding terbalik dengan kesalahan yang terkait dengan estimasi kemampuan. Fungsi informasi tes inilah yang pada akhirnya menggambarkan sejauh mana suatu tes dapat memberikan informasi secara umum.

Aplikasi IRT pada Data Simulasi

Untuk menunjukkan bagaimana IRT dikenakan pada data, maka dalam tulisan ini aplikasi IRT dilakukan dengan menggunakan data simulasi. Data yang digunakan berupa tujuh aitem dengan jumlah

responden 900. Analisis IRT dilakukan tiga kali dengan Program R, berdasarkan model logistik satu, dua, dan tiga parameter. Untuk mengetahui model mana yang tepat digunakan pada data, uji kesesuaian dilakukan terhadap model logistik satu, dua, dan tiga parameter. Hal itu diwujudkan dengan melakukan uji ANOVA berdasarkan ketiga analisis tersebut.

Analisis IRT Model Logistik Satu Parameter

Hasil analisis IRT model logistik satu parameter terhadap tujuh aitem terlihat pada Tabel 1. Dari tabel tersebut diperoleh informasi adanya variasi taraf kesukaran aitem baik bernilai negatif maupun positif. Taraf kesukaran yang bernilai negatif menunjukkan bahwa untuk bisa memperoleh peluang 50% menjawab aitem tersebut dengan benar dibutuhkan kemampuan rendah. Sebaliknya, taraf kesukaran yang bernilai positif menunjukkan bahwa dibutuhkan kemampuan tinggi untuk bisa memperoleh peluang 50% menjawab aitem dengan benar. Pada aitem dengan nilai taraf kesukaran yang negatif, semakin besar angka taraf kesukaran menunjukkan aitem tersebut semakin mudah. Pada aitem dengan nilai taraf kesukaran yang positif, semakin besar angka taraf kesukaran menunjukkan aitem tersebut semakin sukar.

Tabel 1
Hasil Analisis IRT Model Logistik Satu Parameter

| | Taraf Kesukaran | Daya Diskriminasi |
|---------|-----------------|-------------------|
| Aitem 1 | -1,0270726 | 1,159415 |
| Aitem 2 | -1,2668575 | 1,159415 |
| Aitem 3 | -0,6828077 | 1,159415 |
| Aitem 4 | -0,6093514 | 1,159415 |
| Aitem 5 | -0,1432521 | 1,159415 |
| Aitem 6 | 0,7173534 | 1,159415 |
| Aitem 7 | 0,2133976 | 1,159415 |

Dari Tabel 1 diperoleh informasi bahwa taraf kesukaran aitem berkisar antara -1,267 (aitem 2) sampai dengan 0,717 (aitem 6). Aitem dengan taraf kesukaran -1,267 menunjukkan bahwa diperlukan abilitas minimal -1,267 untuk dapat menjawab benar aitem tersebut dengan peluang 50%. Ini berarti aitem tersebut tergolong mudah. Sebaliknya, aitem dengan taraf kesukaran 0,717 tergolong aitem yang sukar karena untuk dapat menjawab benar aitem tersebut dengan peluang 50%, diperlukan abilitas minimal 0,717.

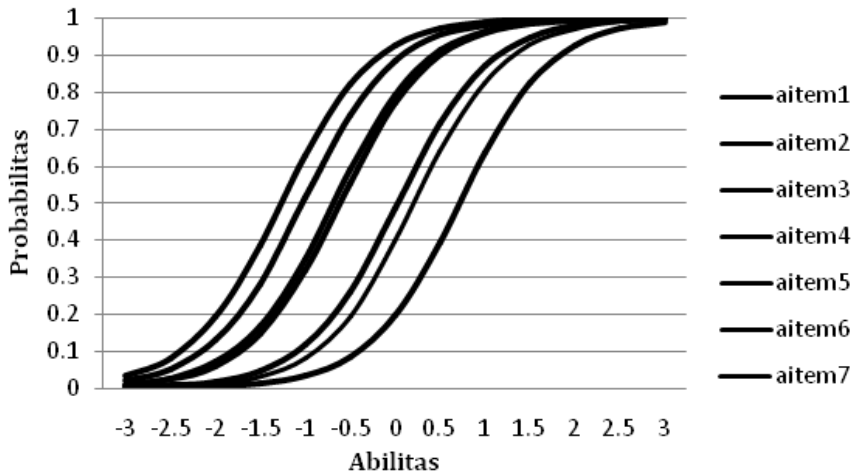
Hambleton *et al.*, (1991) menyatakan bahwa jika abilitas ditransformasikan sehingga rerata menjadi 0 dan deviasi standarnya 1, maka parameter taraf kesukaran akan berada antara -2,0 sampai +2,0. Aitem dengan taraf kesukaran yang mendekati -2,0 menunjukkan aitem yang mudah, sedangkan aitem dengan taraf kesukaran mendekati +2,0 menunjukkan aitem tersebut sukar. Hambleton dan Swaminathan (1985) menyebut parameter taraf kesukaran -2,0 sampai dengan +2,0 sebagai jarak lebar. Dengan jarak tersebut, dapat diperoleh aitem-aitem dengan rentang mudah sampai sukar. Berdasarkan rentang tersebut, ketujuh aitem dalam data simulasi dapat diterima.

Tabel 1 juga menunjukkan bahwa daya diskriminasi yang diperoleh adalah 1,159. Pada model satu parameter, semua aitem ditetapkan memiliki kesetaraan dalam daya diskriminasi. Penyetaraan tersebut ditujukan untuk membebaskan pengaruh karakteristik aitem terhadap estimasi abilitas atau trait laten individu (Sumintono & Widhiarso, 2013). Oleh karena itu, parameter aitem pada model satu parameter difokuskan pada taraf kesukaran saja.

ICC yang terbentuk dari analisis model logistik satu parameter terlihat pada Gambar 3. Dari gambar tersebut dapat terlihat bahwa aitem 2 yang berada pada posisi paling kiri,

merupakan aitem yang termudah. Pada ICC aitem 2, jika ditarik garis vertikal pada probabilitas 0,5, maka abilitas akan menunjuk pada angka -1,267. Sebaliknya, aitem 6 yang berada pada posisi paling kanan merupakan aitem tersulit. Penarikan

garis vertikal pada probabilitas 0,5 akan mengarah pada abilitas 0,717. Kemiringan yang sama pada semua aitem menunjukkan daya diskriminasi yang sama untuk semua aitem.



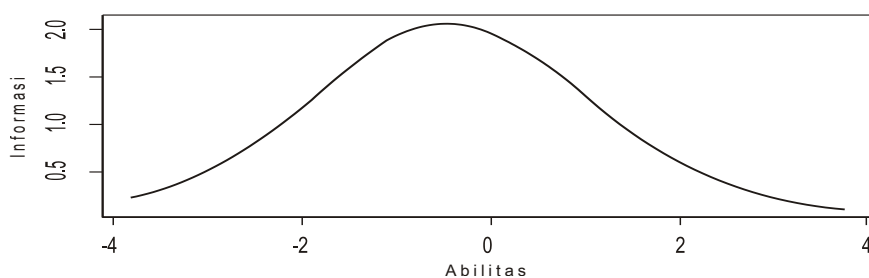
Gambar 3. ICC Berdasarkan Analisis Model Logistik Satu Parameter

Plot fungsi informasi tes berdasar tujuh aitem dengan analisis model logistik satu parameter terlihat pada Gambar 4. Plot tersebut secara umum dapat memberikan informasi bahwa tes dapat mengukur dengan baik khususnya pada kemampuan sedikit di bawah 0. Artinya, tes menghasilkan informasi yang optimal ketika diberikan kepada individu dengan kemampuan yang rendah. Tes yang memusatkan informasi pada level rendah akan lebih sesuai dikenakan pada tes untuk skrining awal atau remedi (Sumintono & Widhiarso, 2013). Tes tersebut kurang cocok jika dipakai untuk seleksi. Untuk keperluan seleksi, diperlukan tes yang bisa memberikan informasi optimal

pada individu-individu dengan kemampuan tinggi.

Analisis IRT Model Logistik Dua Parameter

Hasil analisis IRT model logistik dua parameter terhadap tujuh aitem terlihat pada Tabel 2. Dari tabel tersebut dapat diketahui bahwa aitem yang paling mudah adalah aitem 2 dengan taraf kesukaran -1,089, sedangkan aitem paling sukar adalah aitem 6 dengan taraf kesukaran 1,025. Taraf kesukaran dari ketujuh aitem tersebut berada dalam jarak -2,0 sampai +2,0 seperti pada hasil dari model logistik satu parameter sebelumnya.



Gambar 4. Fungsi Informasi Tes Berdasarkan Analisis Model Satu Parameter

Analisis model logistik dua parameter menghasilkan daya diskriminasi yang berbeda untuk setiap aitem. Pada Tabel 2 dapat diketahui bahwa daya diskriminasi berkisar antara 0,713 – 1,667. Aitem dengan daya diskriminasi yang terendah adalah aitem 6, sedangkan aitem dengan daya diskriminasi tertinggi adalah aitem 3.

Parameter daya diskriminasi bergerak antara 0 sampai 2 (Hambleton, *et al.*, 1991). Eliminasi berdasarkan daya diskriminasi dilakukan terhadap aitem yang memiliki nilai negatif. Aitem dengan daya diskriminasi negatif menunjukkan bahwa probabilitas menjawab benar aitem tersebut justru menurun seiring meningkatnya abilitas responden (Hambleton *et al.*, 1991). Ini berarti ada sesuatu yang salah dengan aitem tersebut. Pada Tabel 2 terlihat bahwa daya diskriminasi pada semua aitem berkisar antara 0-2. Selain itu, pada semua aitem tidak terdapat aitem dengan daya diskriminasi yang bernilai negatif. Oleh karena itu, ketujuh aitem dapat dipakai karena menunjukkan karakteristik yang baik.

ICC yang terbentuk dari analisis model logistik dua parameter terlihat pada Gambar 5. Daya diskriminasi setiap aitem terlihat dari kemiringan ICC dari setiap aitem. Kemiringan yang berbeda pada semua aitem menunjukkan adanya daya diskriminasi

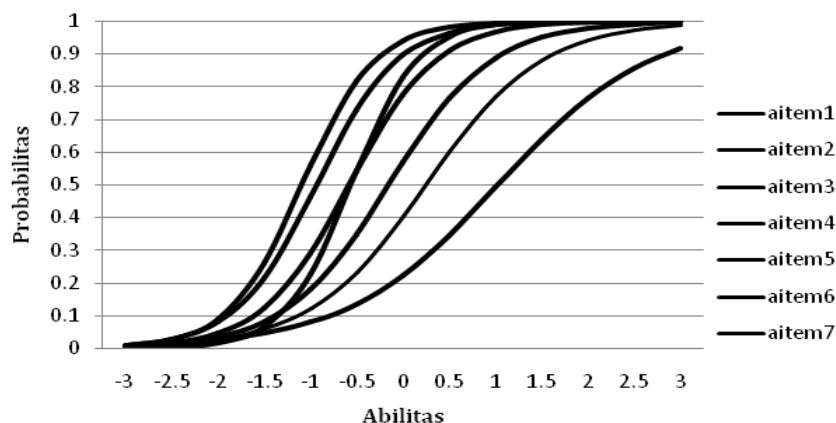
yang bervariasi pada masing-masing aitem. Kemiringan ICC aitem 3 terlihat paling curam di antara aitem-aitem yang lain. Ini berarti aitem tersebut memiliki daya diskriminasi yang paling tinggi. Di sisi lain, kemiringan ICC aitem 6 yang terlihat paling landai mengindikasikan bahwa aitem tersebut memiliki daya diskriminasi paling rendah.

Tabel 2.

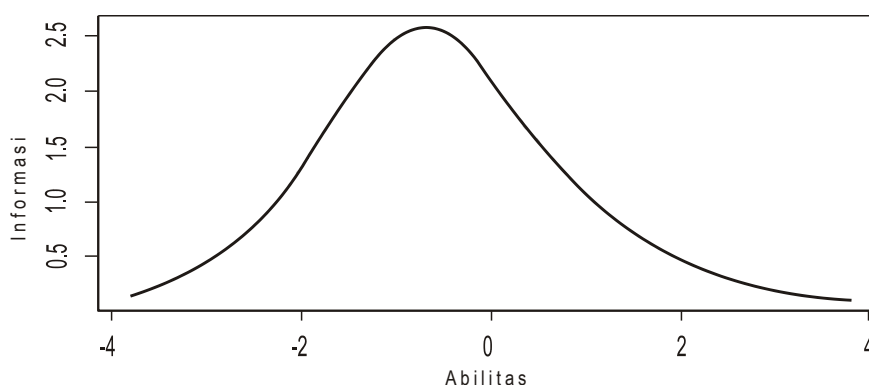
Hasil Analisis IRT Model Logistik 2 Parameter

| | Taraf Kesukaran | Daya Diskriminasi |
|---------|-----------------|-------------------|
| Aitem 1 | -0,9360695 | 1,3557153 |
| Aitem 2 | -1,0888692 | 1,5018927 |
| Aitem 3 | -0,5626501 | 1,6626316 |
| Aitem 4 | -0,5839488 | 1,2524321 |
| Aitem 5 | -0,1547528 | 1,0553781 |
| Aitem 6 | 1,0252841 | 0,7127133 |
| Aitem 7 | 0,2447155 | 0,9395260 |

Plot fungsi informasi tes berdasar tujuh aitem dengan analisis logistik dua parameter terlihat pada Gambar 6. Plot tersebut secara umum dapat memberikan informasi bahwa tes dapat mengukur dengan baik khususnya pada kemampuan sedikit di bawah 0 dengan fungsi informasinya 2,5. Ini berarti tes dapat mendiskriminasi dengan baik pada kemampuan tersebut.



Gambar 5. ICC Berdasarkan Analisis Model 2 Parameter



Gambar 6. Fungsi Informasi Tes Berdasarkan Analisis Model 2 Parameter

Analisis IRT Model Logistik Tiga Parameter

Hasil analisis IRT model logistik tiga parameter terlihat pada Tabel 3. Dari tabel tersebut dapat diketahui bahwa aitem yang paling mudah adalah aitem 2 dengan taraf kesukaran -1,055, sedangkan aitem paling sukar adalah aitem 6 dengan taraf kesukaran 1,246. Untuk parameter daya diskriminasi, aitem dengan daya diskriminasi terendah adalah aitem 7, dengan daya diskriminasi sebesar 0,994. Aitem yang memiliki daya diskriminasi tertinggi adalah aitem 3 dengan daya diskriminasi sebesar 2,534.

Tebakan semu merupakan parameter ketiga yang ada dalam model logistik tiga parameter. Pada Tabel 3 terlihat bahwa tebakan semu berkisar antara 0,000 – 0,286. Nilai tebakan semu yang rendah menunjukkan tingginya peluang aitem tersebut dijawab benar dengan cara ditebak. Sebaliknya,

aitem yang memiliki nilai tebakan semu yang tinggi menunjukkan rendahnya peluang aitem tersebut dijawab secara benar dengan cara ditebak. Dalam seleksi aitem, aitem dengan nilai tebakan semu tinggi harus dieliminasi. Aitem tersebut memungkinkan individu dengan abilitas rendah untuk bisa menjawab secara benar dengan cara menebak jawabannya.

Nilai tebakan semu berkisar antara 0 dan 1. Suatu aitem digolongkan baik jika nilai tebakan semu tidak melebihi $1/k$, dengan k adalah banyaknya pilihan (Hullin dalam Retnawati, 2014). Karena banyaknya pilihan pada data tes adalah empat, maka batas maksimum tebakan semunya adalah $1/4$ atau 0,25. Berdasarkan batas maksimum tersebut, aitem 4 dan 5 layak untuk dieliminasi karena memiliki nilai tebakan semu melebihi batas yang telah ditetapkan.

Tabel 3
Hasil Analisis IRT Model Logistik 3 Parameter

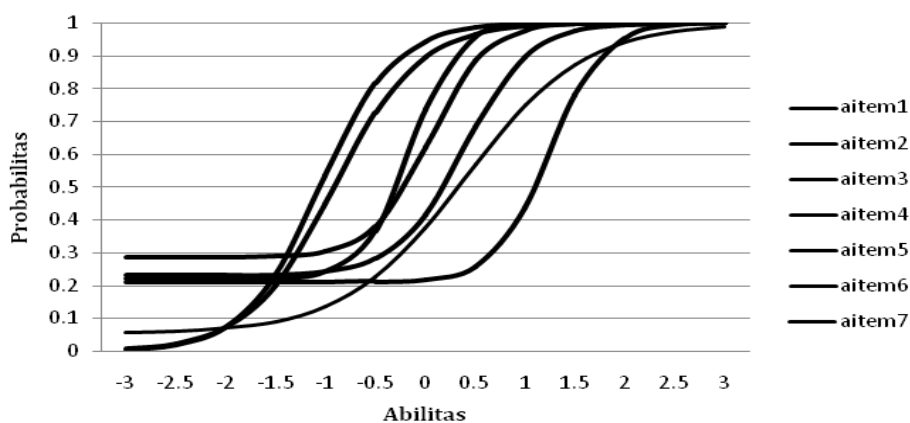
| | Taraf Kesukaran | Daya Diskriminasi | Tebakan Semu |
|---------|-----------------|-------------------|--------------|
| Aitem 1 | -0,915 | 1,395 | 0,001 |
| Aitem 2 | -1,055 | 1,584 | 0,000 |
| Aitem 3 | -0,156 | 2,534 | 0,222 |
| Aitem 4 | 0,035 | 2,085 | 0,286 |
| Aitem 5 | 0,382 | 1,775 | 0,231 |
| Aitem 6 | 1,246 | 2,192 | 0,211 |
| Aitem 7 | 0,389 | 0,994 | 0,053 |

ICC yang terbentuk dari analisis model logistik tiga parameter terlihat pada Gambar 7. Gambar tersebut memberikan informasi bahwa aitem 2 merupakan aitem termudah (posisi paling kiri), sedangkan aitem 6 merupakan aitem tersukar (posisi paling kanan). Untuk parameter daya beda, aitem 3 memiliki daya diskriminasi paling tinggi yang terlihat dengan ICC yang paling tegak dibandingkan ICC aitem-aitem yang lain. Di sisi lain, aitem dengan daya diskriminasi terendah adalah aitem 7. ICC aitem 7 memiliki kemiringan yang paling landai diantara ICC yang lain.

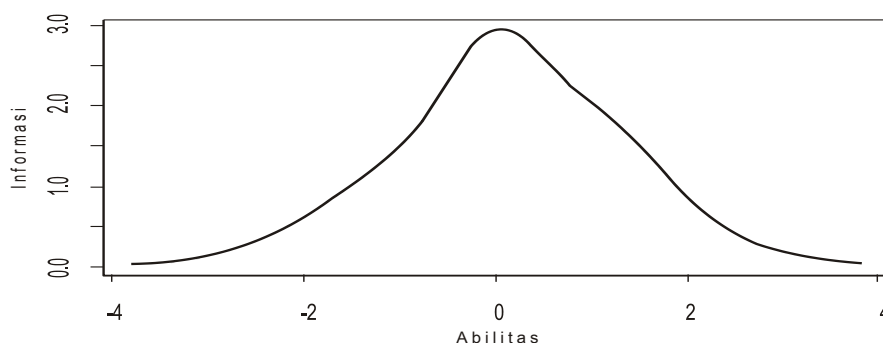
Parameter tebakan semu pada model logistik tiga parameter memungkinkan asimtot bawah berada pada angka lebih dari nol pada ICC. Semakin tinggi asimtot bawah

suatu aitem, semakin tinggi pula peluang aitem tersebut dijawab benar dengan cara ditebak. Pada Gambar 7 terlihat bahwa aitem 2 memiliki asimtot yang paling mendekati nol (peluang untuk menebak paling kecil). Sebaliknya, aitem 4 memiliki asimtot yang paling tinggi. Artinya, aitem tersebut memiliki peluang yang paling tinggi untuk dijawab benar dengan cara ditebak.

Plot fungsi informasi tes berdasar tujuh aitem dengan analisis model logistik tiga parameter terlihat pada Gambar 8. Plot tersebut secara umum memberikan informasi bahwa tes dapat mengukur dengan baik khususnya pada kemampuan sedikit di atas rata-rata. Hasil ini agak berbeda dengan hasil fungsi informasi dari model logistik satu dan dua parameter sebelumnya.



Gambar 7. ICC Berdasarkan Analisis 3 Parameter



Gambar 8. Fungsi Informasi Tes Berdasarkan Analisis Tiga Parameter

Pemilihan Model Analisis

Uji kesesuaian model dengan data merupakan tolak ukur yang dipakai dalam memilih model analisis yang akan dikenakan pada data. Hal tersebut menjadi sesuatu yang penting mengingat analisis yang dilakukan pada akhirnya akan dipergunakan untuk mengestimasi kemampuan individu (Swaminathan, Hambleton, & Rogers, 2007). Pemilihan model analisis yang tidak tepat akan membawa dampak pada timbulnya kesalahan dalam mengestimasi kemampuan individu. Meskipun demikian, perlu untuk diketahui bahwa pada dasarnya tidak ada model yang secara sempurna cocok dengan data (Wiberg, 2004).

Sebelum dilakukan perbandingan model pada data simulasi yang ada, analisis model logistik satu, dua dan tiga parameter dilakukan ulang dengan membuang aitem 4 dan 5 yang memiliki nilai tebakan semu tinggi. Setelah itu, ANOVA dilakukan untuk mengetahui manakah di antara model logistik satu dan dua parameter yang lebih sesuai digunakan pada data. Hasil ANOVA tersebut terlihat pada Tabel 4. Harga $p < 0,001$ menunjukkan bahwa terdapat perbedaan antara model satu dan dua parameter.

Akaike Information Criteria (AIC) digunakan untuk mengatasi permasalahan pemilihan model. AIC diformulasikan untuk memilih model 'perkiraan terbaik' di antara beberapa model pengukuran dengan jumlah parameter yang berbeda, berdasarkan kriteria statistik yang cocok (Everitt & Howell, 2005). Model yang terbaik adalah

model dengan skor AIC paling rendah (Snipes & Taylor, 2014). Pada Tabel 4 terlihat bahwa nilai AIC model logistik dua parameter (5363,58) menunjukkan hasil yang lebih rendah dibandingkan dengan model logistik satu parameter (5374,16). Ini berarti model logistik dua parameter memiliki kesesuaian terhadap data yang lebih baik dibandingkan dengan model logistik satu parameter.

ANOVA kedua dilakukan untuk membandingkan model logistik dua dan tiga parameter. Hasil ANOVA tersebut terlihat pada Tabel 5. Harga $p > 0,05$ pada tabel tersebut menunjukkan bahwa tidak terdapat perbedaan kesesuaian data antara model logistik dua dan tiga parameter.

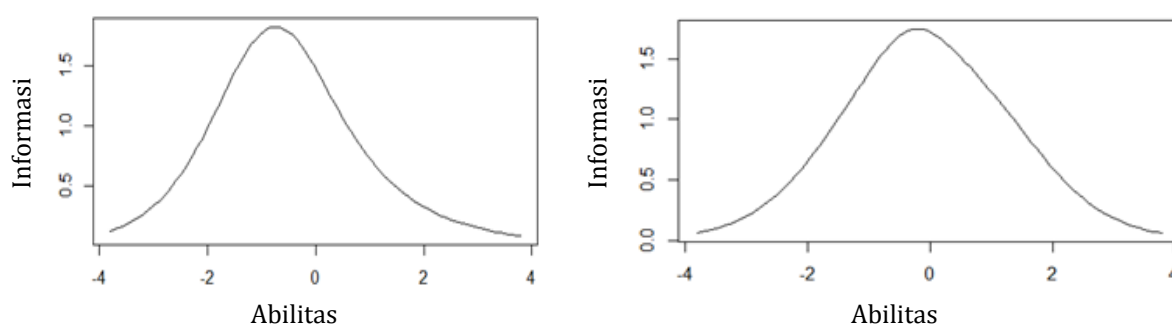
Plot fungsi informasi model logistik dua dan tiga parameter berdasarkan aitem yang terseleksi terlihat pada Gambar 9. Secara umum keduanya plot tersebut nampak serupa. Namun demikian plot fungsi informasi model logistik dua parameter terlihat lebih memiliki informasi yang tinggi pada kemampuan responden di bawah rata-rata. Ini berarti tes tersebut dapat menjadi pilihan jika tes dimaksudkan untuk fungsi skrining misalnya. Plot fungsi informasi model tiga parameter cenderung memberikan informasi optimal pada individu dengan abilitas 0. Apabila tes tersebut ditujukan untuk fungsi seleksi, pengembang tes perlu menambahkan aitem-aitem lain yang memberikan informasi tinggi pada abilitas tinggi (Furr, 2011).

Tabel 4.
Hasil ANOVA Model 1 dan 2 Parameter

| | AIC | Logaritma Kebolehhadian | Rasio Kebolehhadian | Derajat kebebasan | Harga p |
|------------|---------|-------------------------|---------------------|-------------------|---------|
| Model 1 PL | 5374,16 | -2681,08 | | | |
| Model 2 PL | 5363,58 | -2671,79 | 18,58 | 4 | <0.001 |

Tabel 5.
Hasil ANOVA Model 2 dan 3 Parameter

| | AIC | Logaritma Kebolehjadian | Rasio Kebolehjadian | Derajat kebebasan | Harga p |
|------------|---------|----------------------------|------------------------|----------------------|---------|
| Model 2 PL | 5363,58 | -2671,79 | | | |
| Model 3 PL | 5369,36 | -2669,68 | 4,22 | 5 | 0.519 |



Gambar 9. Perbandingan Fungsi Informasi Tes Berdasarkan Analisis 2 & 3 Parameter dengan Aitem Terseleksi

Dalam analisis IRT, fungsi informasi dan simpangan baku pengukuran layak dipertimbangkan dalam menentukan model yang dipilih (Ridho, 2007). Fungsi informasi menunjukkan sejauh mana masing-masing model mampu memberikan informasi (Veerkamp & Berger, 1999). Semakin tinggi puncak dari fungsi informasi, semakin tinggi pula informasi yang bisa diberikan oleh suatu model. Hal ini berkebalikan dengan simpangan baku pengukuran yang diharapkan rendah. Ridho (2007) menggunakan kedua kriteria tersebut dalam memilih model dalam penelitiannya. Dalam penelitian tersebut model dua parameter dipilih sebagai model dibandingkan model satu dan tiga parameter, berdasarkan fungsi informasi tertinggi serta simpangan baku pengukuran terendah.

Penutup

IRT sangat bermanfaat dalam analisis aitem suatu tes. Melalui IRT, indeks parameter-parameter aitem dapat diketahui dengan mudah. Indeks tersebut menjadi dasar

dalam melakukan seleksi aitem. Selain itu, fungsi informasi dapat memberikan pertimbangan bagaimana sebaiknya tes digunakan. Keberadaan IRT juga didukung dengan adanya Program R yang dapat diunduh secara bebas tanpa berbayar. Oleh karena itu, penggunaan IRT dalam pengembangan alat ukur psikologi perlu untuk ditingkatkan.

Daftar Pustaka

- Azwar, S. & Ridho, A. (2013). Abilitas komposit pada tes potensi. *Jurnal Psikologi*, 40(2), 127-142.
- Everitt, B., & Howell, D. C. (Eds.). (2005). *Encyclopedia of statistics in behavioral science*. Hoboken, N.J: John Wiley & Sons.
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. London: SAGE.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory, principles and*

- applications*. New York: Springer Science+Business Media.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications, Inc.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates.
- Molenaar, I. W. (1995). Some background for item response theory and the rasch model. Dalam G. H. Fischer, & I. W. Molenaar (Eds). *Rasch models*. New York: Springer-Verlag.
- Retnawati, H. (2014). *Teori respons butir dan penerapannya*. Yogyakarta: Parama Publishing.
- Ridho, A. (2007). Karakteristik psikometrik tes berdasarkan pendekatan teori tes klasik dan teori respon aitem. *Jurnal Psikologi INSAN*, 2(2), 1-27.
- Snipes, M., & Taylor, D. C. (2014). Model selection and Akaike information criteria: An example from wine ratings and prices. *Wine Economics and Policy*, 3(1), 3–9. doi: 10.1016/j.wep. 2014.03.001
- Sumintono, B, & Widhiarso, W. (2013). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial*. Cimahi: Trim Komunikata Publishing House.
- Swaminathan, H., Hambleton, R. K, & Rogers, H. J. (2007). Assessing the fit of item response models. *Handbook of Statistics*, 26, 683-718.
- Veerkamp, W. J. & Berger, M. P. (1999). Optimal item discrimination and maximum information for logistic IRT models. *Applied Psychological Measurement*, 23(1), 31-40.
- Wiberg, M. (2004). Classical test theory vs. item response theory. *Umea*, 10(5), 1–27.
- Widhiarso, W. (2016). Eksplorasi gaya respons ekstrem dalam mengisi kuesioner. *Jurnal Psikologi*, 43(1), 16-29. doi:10.22146/jpsi.8703