

## Analisis Rasch Model Indonesia the International Personality Item Pool-Big Five Factor Markers (IPIP-BFM-50)

Timothy Ryan

Fakultas Psikologi, Universitas Surabaya, Jl. Raya Rungkut, Kali Rungkut, Rungkut, Surabaya, Jawa Timur

Stevanny Angella

Fakultas Psikologi, Universitas Surabaya, Jl. Raya Rungkut, Kali Rungkut, Rungkut, Surabaya, Jawa Timur

Roy Surya\*

Fakultas Psikologi, Universitas Surabaya, Jl. Raya Rungkut, Kali Rungkut, Rungkut, Surabaya, Jawa Timur

\*E-mail : roysuroya99@gmail.com

### Abstract

Various studies have been conducted to determine human personality, then there are various theories and personality scales. IPIP-BFM-50 is open source and has been used around the world. The aims of this research is to validate IPIP-BFM-50 with rasch model. The sample used was 359 students from University of Surabaya, with an age range 17-23 years. This research used simple random sampling. The outfit-infit statistic results vary from 0.5-1.5 (except C-28), person-item reliability is above 0.7, the passable dimesionality measure.. There are several items that indicate bias based on DIF measurement: EM-44, EM-14, EM-29, EM-9, E-6, E41, C-38, A-32, and I-30. There are some items on the dimensions of emotional stability, extraversion, conscientiousness, agreeableness, and intellect with less varied levels of DIF difficulty. Researcher proposed choice to rework this instrument in the terms for fixed its item bias and wider its variability (DIF difficulties hierarchy). This studies also implies on similar studies to use both CTT and rasch model simultaneously for gain richer psychometrics information.

**Keywords:** IPIP-BFM-50; Personality; Rasch Model

### Abstrak

Berbagai penelitian telah dilakukan untuk mengetahui kepribadian manusia, yang kemudian muncul berbagai teori dan alat ukur kepribadian. IPIP-BFM-50 merupakan salah satu alat ukur kepribadian open source dan telah banyak digunakan di seluruh dunia. Penelitian ini memiliki tujuan melengkapi studi validasi IPIP BFM-50 sebelumnya, dengan menggunakan rasch model. Sampel yang digunakan yaitu 359 mahasiswa Universitas Surabaya, dengan usia 17-23 tahun. Pengambilan sampel dilakukan dengan teknik simple random sampling. Diperoleh hasil outfit-infit seluruh butir yang bervariasi pada 0.5 – 1.5 (kecuali C-28), reliabilitas person-item pada semua dimensi dari alat ukur ini diatas 0.7, hasil uji dimensionalitas yang cukup memuaskan. Terdapat beberapa butir yang terindikasi bias yaitu butir EM-44, EM-14, EM-29, EM-9, E-6, E41, C-38, A-32, dan I-30. Beberapa butir pada dimensi emotional stability, extraversion, dan conscientiousness, agreeableness, dan intellect memiliki tingkat kesulitannya kurang bervariasi. Peneliti menyarankan untuk melakukan revisi dari instrumen ini apabila berkeinginan untuk memperbaiki butir yang diindikasikan mengalami bias dan memperluas variasi tingkat kesulitan butir. Studi ini juga berimplikasi pada studi serupa berikutnya terkait validasi dan adaptasi untuk menggunakan kedua paradigma CTT dan rasch model sehingga mendapatkan informasi psikometris yang lebih banyak dan sempurna.

**Kata kunci:** IPIP-BFM-50; Kepribadian; Rasch Model

Copyright © 2022. Timothy Ryan, Stevanny Angella, Roy Surya

Submitted: 2021-09-09

Revised: 2021-11-11

Accepted: 2022-03-31

Published: 2022-04-01

---

## Pendahuluan

Kepribadian merupakan salah satu konsep yang sangat kompleks untuk dapat dideskripsikan secara sederhana. Banyak teori mengenai kepribadian yang terus berkembang dan bertambah variasinya, semisal konsep kepribadian *big five* dan kepribadian *big-six*. Telah banyak penelitian dilakukan untuk mengembangkan instrumentasi untuk mengukur, mengevaluasi atau menilai kepribadian manusia yang kompleks tersebut (Schultz & Schultz, 2009). *Big Five Personality* merupakan salah satu teori kepribadian yang sedang berkembang hingga saat ini. Terdapat dua model yang menjelaskan teori ini yaitu *big five* yang dikembangkan oleh Goldberg dan *five factor* model yang dikembangkan oleh Costa dan McCrae (Akhtar & Azwar, 2019).

IPIP (*International Personality Item Pool*) merupakan salah satu sarana yang dipergunakan untuk pengukuran kepribadian, dikembangkan dengan kolaborasi skala internasional untuk pengembangan inventori kepribadiannya, hingga saat ini terdapat total 3.320 item yang berisikan berbagai macam skala di dalamnya seperti BFQ-C, (Gaio, 2012), Neo-PI-R, (Costa, 2018), Mini IPIP, (Donnellan dkk., 2006), IPIP-Neo, (Johnson, 2014), IPIP HEXACO, (Babarović & Šverko, 2013), dan masih banyak lagi. IPIP pertama kali diperkenalkan pada tahun 1996, awalnya alat ukur ini berjumlah 1252, kemudian berkembang menjadi lebih dari 2000 dan terdapat item baru yang ditambahkan setiap tahunnya (Goldberg dkk., 2006). Alat ukur ini dikembangkan oleh *the Groningen personality team*, mereka mengembangkan lebih dari 1000 butir dalam bahasa Belanda, kemudian Goldberg bergabung dan menerjemahkannya ke dalam bahasa Inggris. Terbentuklah 1.252 butir bahasa Inggris dari hasil kolaborasi mereka (Maples dkk., 2014).

Hal yang membedakan antara alat ukur IPIP dengan alat ukur lain adalah memiliki cara unik, dalam perkembangannya IPIP-BFM mengadopsi format butir yang lebih kontekstual dan lebih panjang dari trait adjektif (Lim & Ployhart, 2006). IPIP-BFM pada awalnya memiliki skala dengan dua macam versi yang populer, berisikan 50 butir dan 100 butir, yang telah diadaptasi ke lebih dari 10 negara, beberapa diantaranya seperti Polandia, (Strus dkk., 2014), China, (Zheng dkk., 2008), Skotlandia, (Gow dkk., 2005), Argentina, (Cupani & Lorenzo-Seva, 2016), Portugal, (Oliveira, 2019), serta Indonesia (Akhtar & Azwar, 2019). Dimensi kepribadian dalam IPIP-BFM meliputi *Extraversion*, *Agreeableness*, *Conscientiousness*, *Emotional stability* dan *Intellect*. Hasil dari penelitian sebelumnya milik (Akhtar & Azwar, 2019a) yang telah mengadaptasi instrumen BFM-50 pada populasi di Indonesia menunjukkan jika alat ukur ini telah memiliki sifat psikometri yang memuaskan serta dapat dipahami oleh responden dengan masing-masing koefisien reliabilitasnya sebesar, *emotional stability*  $\alpha = 0.86$ , *extraversion*  $\alpha = 0.87$ , *conscientiousness*  $\alpha = 0.811$ , *agreeableness*  $\alpha = 0.83$ , *intellect*  $\alpha = 0.768$  serta terbaginya semua butir pada kelompok-kelompok yang pada semestinya, dengan catatan pada butir 145 di dimensi *intellect* yang memiliki nilai *faktor loading* rendah, 0.34.



Secara definitif untuk menjelaskan secara singkat bagian-bagian dimensinya, Individu dengan *extraversion* yang tinggi dinilai memiliki kecenderungan untuk lebih merasa bahagia, lebih positif serta termotivasi dalam melakukan interaksi sosial; sementara Individu yang memiliki *Agreeableness* tinggi dapat diamati sebagai orang yang terlihat bersahaja, kooperatif, sederhana, percaya dan bersimpati; berikutnya Individu dengan *Conscientiousness*, merupakan orang yang diamati memiliki perilaku yang dapat diandalkan oleh sesamanya, tepat waktu, teratur dan disiplin (Weiten, 2013). *Emotional stability* sering dianggap sebagai kebalikan dari *Neuroticism* dan merupakan abilitas seseorang untuk menahan stress yang dialaminya, Individu yang memiliki *emotional stability* rendah atau emosi yang tidak stabil cenderung mudah gelisah, murung, sedih, dan penuh ketegangan (Sari, 2016). Menurut penelitian (DeYoung, dkk., 2009; Nusbaum & Silvia, 2011 dalam (Ingram dkk., 2013)) definisi dari *Intellect* berkaitan dengan segala aktivitas otak yang mengukur *working memory*, serta kaitannya dengan *fluid intelligence*. *Intellect* merupakan konsep yang mengakar pada penggambaran kecerdasan dan gaya intelektual seseorang. Meskipun, dimensi *Intellect* dan *Openness* sama-sama mendasari domain mengenai keterbukaan, keduanya dinilai berbeda melalui analisis faktor pada beberapa studi (DeYoung, Peterson, & Quilty, 2007; Johnson, 1994 dalam (Ingram dkk., 2013)).

Peneliti melihat terdapat beberapa hal yang belum dapat dijangkau dari hasil validasi sebelumnya yang menggunakan basis analisis statistik CTT (*classical test theory*) terutama pada tingkat kesulitan butir, komparasi tingkat kesulitan butir serta kaitannya dengan responden, reliabilitas responden, dan validitas skala respon dari butir-butir yang ada. Lebih lanjut secara sederhana, CTT dapat digambarkan sebagai analisis yang berorientasi pada informasi psikometri dari seluruh skala, sedangkan IRT (*item response theory*) memiliki abilitas lain dalam kemampuannya untuk meninjau informasi psikometri hingga tahap butir serta skala secara keseluruhan (Mallinckrodt dkk., 2016). Pendekatan IRT sendiri mempertimbangkan pola skor yang ada pada saat menyimpulkan atribut skor laten, artinya IRT dapat memperkirakan perubahan kecil yang mungkin kurang diperhatikan jika peneliti menggunakan CTT (Jabrayilov dkk., 2016). Beberapa studi serupa terkait validasi alat ukur untuk mencari informasi psikometri yang ada, digunakan pendekatan statistik CTT dengan confirmatory factor analysis dan pendekatan statistik IRT dengan rasch model secara bersamaan, (Ashley dkk., 2013; Lin dkk., 2019; Sakib dkk., 2020; Sondergeld & Johnson, 2014; Taasobshirazi dkk., 2015; Yan dkk., 2020). Penggunaan analisis *rasch model* mulai sering digunakan karena dapat memungkinkan peneliti untuk dapat menelaah butir sesuai dengan permodelan yang berlandaskan asumsi rasch model, memungkinkan estimasi yang tepat dari skor sebenarnya, membantu mengurangi bias jenis kelamin, membantu untuk melihat tingkat kesulitan dari butir-butir yang ada secara lebih spesifik dan detail, dengan demikian peneliti dapat memperoleh lebih banyak informasi tentang proses pengukuran (Mallinckrodt dkk., 2016; Rusch dkk., 2012) (Mallinckrodt dkk., 2016; Rusch dkk., 2017).

---

*Rasch* model sendiri merupakan metode analisis data yang dikembangkan oleh George Rasch, (Rasch, 1980 dalam (Brandt dkk., 2015)), model ini pada awalnya digunakan untuk keperluan di lingkungan pendidikan namun saat ini juga dikembangkan ke dalam beberapa penelitian ilmiah di area lainnya, (Hagquist dkk., 2009). Keistimewaan analisis apabila menggunakan model ini adalah ketidakbergantungannya analisis pada karakteristik data yang diambil secara penuh sehingga pengukuran yang dihasilkan akan lebih objektif seperti halnya membuat jangkar terkait standar parameter yang lebih tinggi (sulit) atau rendah (mudah) terkait dengan abilitas seseorang dan butir, beberapa indikator lain yang dapat diamati dari analisis ini seperti *infit-outfit*, reliabilitas *person-item*, separasi *person-item*, *point measure correlation*, hasil pengukuran logit, *wright map*, skala peringkat, hingga grafik ICC (Bond & Fox, 2015). Skala logit yang dihasilkan dalam analisis *rasch* model nantinya merupakan hasil dari nilai *odd ratio* yang kemudian ditransformasikan dengan fungsi logaritma natural, selain itu skala yang dihasilkan akan memunculkan nilai yang berdependensi dengan pola responnya sehingga *rasch model* bisa memberikan hasil pengukuran yang lebih independen (Sumintono, 2014).

Alat ukur IPIP ini sebelumnya telah diadaptasi kedalam bahasa Indonesia oleh (Akhtar & Azwar, 2019) dengan menggunakan analisis statistik CTT. Saat ini belum ada penelitian serupa yang menggunakan *rasch* model. Lebih lanjut penelitian ini dilakukan dengan tujuan untuk melengkapi kelemahan proses validasi sebelumnya serta memperkaya data informasi psikometris yang belum dapat dijangkau, dengan menggunakan analisis *rasch model*.

## Metode

### Partisipan Penelitian

Pada penelitian ini sampel berjumlah 359 mahasiswa Universitas Surabaya, terdiri atas 266 perempuan (74.09%) dan 93 laki-laki (25.9%), sampel merupakan mahasiswa berusia 17-23 tahun dengan rentang usia 17[D1]-23 tahun. Teknik pengambilan sampel yang digunakan adalah *simple random sampling*. Dalam teknik *simple random sampling* setiap individu memiliki kesempatan yang sama untuk dipilih menjadi sampel. Kelebihan dari teknik ini adalah tingginya validitas internal dan eksternal serta kemudahan dalam menganalisis data. Sedangkan kelemahan dari teknik ini adalah memerlukan kerangka *sampling* dan biaya yang besar (Acharya dkk., 2013), meskipun dalam analisis *rasch* tidak memerlukan parameter sampel yang begitu besar.

Menurut (Linacre, 1994) dalam analisis *rasch* dengan model *polythomus* data, sampel sebanyak minimal 50 sebenarnya dapat dikatakan cukup untuk memberikan informasi terhadap analisis instrumen terkait dengan kalibrasi butir  $\pm 1$  logit serta tingkat kepercayaan 95%, namun apabila ukuran sampel semakin banyak tentu akan semakin baik adanya. Pada studi milik (Smith dkk., 2008), dengan asumsi *rasch model* untuk *polytomous data* ditemukan jika responden dengan jumlah yang kurang dari 50 memperlihatkan banyak *misfit* pada analisis butir daripada 100 responden, berikutnya jika melibatkan lebih



banyak dari 100 responden hingga lebih dari 200 akan membuat pengukuran lebih *robust* dengan presisi yang lebih baik.

### **Instrumen Penelitian**

Alat ukur yang dipergunakan adalah *International Personality Item Pool (IPIP)-BFM 50* yang sudah diadaptasi sebelumnya oleh (Akhtar & Azwar, 2019), sehingga peneliti terlebih dahulu peneliti mengurus perizinan sebelum menggunakannya. IPIP-BFM 50 memiliki komposisi 10 butir pada setiap dimensinya, menggunakan skala *likert* mulai dari skala 1 yang menggambarkan sangat tidak sesuai hingga 5 yang menggambarkan sangat sesuai, dengan demikian sebelum memulai analisis, peneliti akan mengasumsikan semua butir adalah *favorable* dengan upaya membalik hasil skoring pada butir-butir *unfavorable*. Uji validitas isi dari alat ukur ini telah dilakukan oleh (Akhtar & Azwar, 2019) dengan menggunakan metode Aiken's V, terbukti bahwa seluruh item memiliki nilai indeks diatas 0.69. Lalu diketahui juga bahwa alat ukur ini memiliki reliabilitas yang diatas 0.7, yang menandakan bahwa setiap dimensi memiliki konsistensi dalam pengukuran yang baik.

### **Teknik Analisis Data**

Analisis data yang digunakan untuk uji validitas dan reliabilitas dalam penelitian ini menggunakan pendekatan *rasch model*. Analisis data akan berfokus pada nilai *infit-outfit meansquare* untuk melihat ketepatan data yang diperoleh dengan model, reliabilitas *person-items* untuk melihat kemungkinan adanya *outliers* atau responden yang tidak fit dengan model, hasil pengukuran logit untuk melihat tingkat kesulitan butir, *wright map* untuk meninjau relasi *person* dan *item*, menggunakan Winstep 5.2.2 (Linacre, 2022) serta sebagai analisis tambahan, *Differential item functioning (DIFF)* untuk melihat adanya bias dari butir-butir yang ada ditinjau dari jenis kelamin (laki-laki dan perempuan) menggunakan DIFAS 5.0, (Penfield, 2013).

Analisis dimensionalitas dilakukan secara terpisah pada masing-masing dimensi, sesuai hasil confirmatory factor analysis pada penelitian sebelumnya, mulai dari *emotional stability*, *extraversion*, *conscientiousness*, *agreeableness*, hingga *intellect*, karena asumsi dimensionalitas pada analisis *rasch* yang menurut (Bond & Fox, 2015) merupakan asumsi yang unidimensional.

### **Hasil**

Pada uji unidimensionalitas diketahui 4 dari 5 dimensi menghasilkan unidimensionalitas nilai presentase diatas 40%. Pada dimensi *emotional stability (EM)* = 46.2%; dimensi *extraversion (E)* = 47.7%; dimensi *agreeableness (A)* = 44.5%; dimensi *intellect (I)* = 41%. Berdasarkan (Linacre, 2011), nilai uji dimensionalitas yang berada lebih dari 40% menunjukkan kesesuaian dari butir-butir didalamnya dengan konstruksi definisi dimensi yang dimaksud. Keempat dimensi tersebut memiliki butir-butir yang mengukur apa yang memang benar ingin diukur, dan semakin besar nilai presentasi menandakan semakin berkurangnya pengotor yang mengganggu *infit-outfit* statistik butir. Uji dimensionalitas pada dimensi *Conscientiousness (C)* = 36.6%, hal ini secara umum dimungkinkan terjadi

karena dimensi ini memiliki butir-butir yang terlalu *underfit/overfit* atau *noise* dari analisis statistik *infit-outfit* beberapa responden yang tidak sempurna berada pada permodelan *rasch* yang menurut (Bond & Fox, 2015) memiliki batas *mnsq* ( $0.5 < x < 1.5$ ), meskipun demikian, analisis *infit-outfit* responden menunjukkan mayoritas responden telah memiliki nilai *fit* yang telah mengikuti permodelan *rasch*.

Reliabilitas dari alat ukur ini dinilai memuaskan, dengan *Item reliability* yang dimiliki oleh kelima dimensi berada pada rentang 0.83-0.99. Jika dilihat dari *person reliability*-nya, diketahui bahwa skor berada pada rentang 0.75-0.83, baik pada reliabilitas *item* dan *person* sama-sama menunjukkan nilai diatas batas yang dianjurkan ( $x < 0.7$ ), hal ini menurut (Linacre, 2011) menunjukkan jika sampel yang digunakan cukup besar untuk dianalisis serta menandakan cukup tingginya peluang orang/butir yang diukur dengan abilitas tinggi sebenarnya memang benar memiliki abilitas yang tinggi, sebaliknya demikian juga dengan orang/butir dengan abilitas rendah. Jika diamati, dimensi C, A, dan I memiliki nilai reliabilitas yang lebih tinggi dibanding dengan dimensi EM dan E, hal ini mungkin untuk terjadi dikarenakan butir-butir dalam dimensi C, A, dan I lebih bervariasi tingkat kesulitannya. Indeks separasi menandakan persebaran butir/orang pada mistar logit pada *wright map*, semakin tinggi nilai separasi orang menandakan instrumen memiliki sensitifitas yang tinggi untuk membedakan orang dengan kemampuan rendah atau tinggi sementara itu semakin tinggi separasi butir menandakan sampel dari responden semakin beragam, (Linacre, 2011). Indeks separasi *person* pada kelima dimensi berada para nilai 2.21 - 1.74, menandakan jika butir-butir pada kelima dimensi tersebut dinilai kurang bervariasi untuk mendiskriminasikan keberagaman sampel yang ada. Meskipun demikian, indeks separasi butir EM, E, dan C dengan rentang nilai 1.81-2.24 menandakan jika sampel yang menggambarkan responden dengan abilitas di masing-masing dimensi EM, E dan C dinilai kurang baik dalam menggambarkan kesulitan butir-butir pada dimensi tersebut, sementara itu orang-orang dengan abilitas pada dimensi A dan I, dinilai lebih baik dan beragam untuk menggambarkan berbagai tingkat kesulitan butir-butir pada dimensi A dan I. Nilai yang disarankan untuk batas indeks separasi butir dan orang berada pada diatas 3 untuk hasil yang lebih baik, (Linacre, 2011).

**Tabel 1**

Analisis Umum Ke-5 Dimensi

Dimensi	Reliabilitas Item	Reliabilitas Person	Dimensionalitas	Indek Item	Separasi	Indeks Separasi Person
<i>Emotional Stability</i> (EM)	0.83	0.83	46.2%	2.24		2.21
<i>Extraversion</i> (E)	0.83	0.83	47.7%	2.24		2.22
<i>Conscientiousness</i> (C)	0.96	0.75	36.6%	1.81		1.74
<i>Agreeableness</i> (A)	0.99	0.76	44.5%	8.58		1.77
<i>Intelect</i> (I)	0.98	0.76	41%	7.42		1.79



Pada analisis validitas skala peringkat untuk setiap dimensi, ditemukan nilai yang memuaskan pada kolom *andrich treshold* dan *category measure*. Hal ini menandakan skala peringkat yang digunakan tidak membingungkan responden dan dengan baik mampu membedakan responden dengan kemampuan rendah hingga tinggi. Asumsinya adalah seharusnya orang dengan kemampuan rendah memiliki probabilitas yang tinggi dalam memilih skala peringkat 1 dan akan turun seiring dengan meningkatnya kemampuan orang lain dengan kemampuan yang lebih tinggi, begitu pun juga sebaliknya pada skala peringkat 5, (Bond & Fox, 2015).

### **Emotional Stability**

*Point measure correlation* pada dimensi ini berkisar pada nilai 0.52-0.72 dan tidak ditemukan nilai yang negatif. Hal tersebut menandakan bahwa butir-butir yang ada di dalam dimensi ini mampu dengan baik mendiskriminasikan orang dengan kemampuan rendah dan tinggi. Batas nilai ideal yang dianjurkan untuk *point measure correlation* adalah lebih dari 0.3, (Linacre, 2011). Apabila dilihat dari nilai logit, diketahui bahwa butir yang paling sulit untuk mendapatkan persetujuan oleh responden adalah butir EM-14 dengan skor 0.75, sementara EM-9 dan EM-49 merupakan butir yang paling mudah mendapatkan persetujuan responden. Nilai *infit-outfit meansquare* pada dimensi *emotional stability*, menunjukkan jika butir-butir di dalam dimensi ini fit atau sesuai dengan model, dengan batas yang menurut (Bond & Fox, 2015) berkisar antara 0.5-1.5.

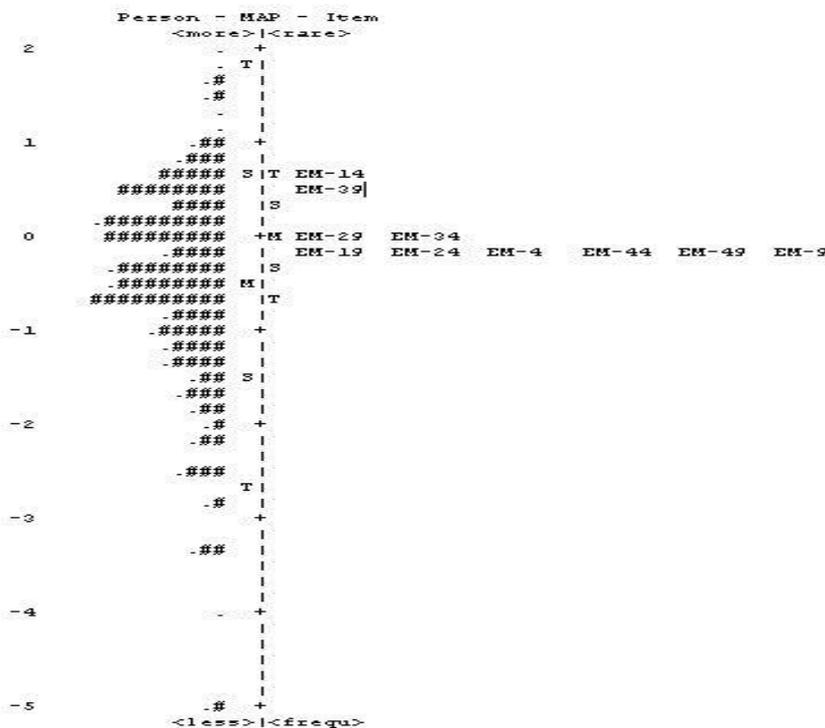
**Tabel 2**  
Analisis Dimensi *Emotional Stability*

	<i>Measure</i>	<i>S.E.Measure</i>	<i>Infit MS</i>	<i>Outfit MS</i>	<i>Point measure corr.</i>
EM14	0.75	0.07	1.01	0.95	0.63
EM39	0.42	0.06	0.86	0.86	0.69
EM34	0.06	0.06	0.91	0.90	0.71
EM29	-0.08	0.06	0.89	0.88	0.69
EM4	-0.10	0.06	0.92	0.89	0.72
EM24	-0.16	0.06	1.22	1.27	0.58
EM44	-0.17	0.06	0.90	0.89	0.69
EM19	-0.22	0.06	1.16	1.27	0.57
EM49	-0.25	0.06	1.02	0.99	0.71
EM9	-0.25	0.06	1.12	1.35	0.52

Pada analisis persebaran butir, didapati terlalu banyak butir dengan nilai logit yang kurang bervariasi atau memiliki perbedaan nilai yang sangat dekat seperti butir EM-34=(-0.2196), EM-29=(-0.2462), EM-4=(-0.2462), EM-24=(-0.16), EM-44=(-0.17), dan EM-19=(-0.22). Lebih lanjut butir EM-49=(-0.25) dan EM-9=(-0.25) memiliki nilai logit yang sama. Butir yang memiliki nilai logit hampir sama atau sama, nantinya akan sulit untuk mendiskriminasikan abilitas individu dengan baik, mengingat dasar persamaan rasch

model yang tidak menyertakan fungsi linear dengan  $\alpha (\theta - \delta_j)$ , (de Ayala, 2009). Butir-butir pada dimensi *emotional stability* akan sangat efektif untuk mengukur individu dengan abilitas menengah keatas dibandingkan dengan abilitas yang lebih rendah, oleh karena itu pada dimensi ini orang-orang dengan tingkat *emotional stability* yang rendah, kurang mampu untuk bisa atau dengan kata lain butir-butir pada dimensi ini akan sangat cocok untuk diberikan kepada orang dengan abilitas *emotional stability* yang tinggi. Butir dengan nilai yang berdekatan juga memberikan inefisiensi dalam tes dikarenakan memberikan informasi abilitas yang hampir sama.

**Gambar 1**  
Wright Map Dimensi Emotional Stability



**Extraversion**

*Point measure correlation* yang dianjurkan beradalebih dari nilai 0.3, (Linacre, 2011), pada dimensi ini nilainya berkisar pada 0.48-0.74 dan tidak ditemukan adanya nilai yang negatif. Hal ini menandakan bahwa butir-butir yang ada mampu dengan baik untuk membedakan orang-orang dengan abilitas yang rendah dan tinggi. Apabila dilihat dari nilai logit, diketahui bahwa butir yang cenderung sulit untuk mendapatkan persetujuan oleh responden adalah butir E-16 = 0.80 dan E-46 = 0.64, sementara itu butir E-21 = (-0.07) merupakan butir yang cenderung paling mudah mendapatkan persetujuan responden, terutama bagi responden dengan abilitas yang tinggi. Nilai *infit-outfit meansquare* pada dimensi *emotional stability*, menunjukkan jika butir-butir di dalam dimensi ini fit atau sesuai dengan model, dengan batas yang menurut (Bond & Fox, 2015) berkisar antara 0.5-1.5.

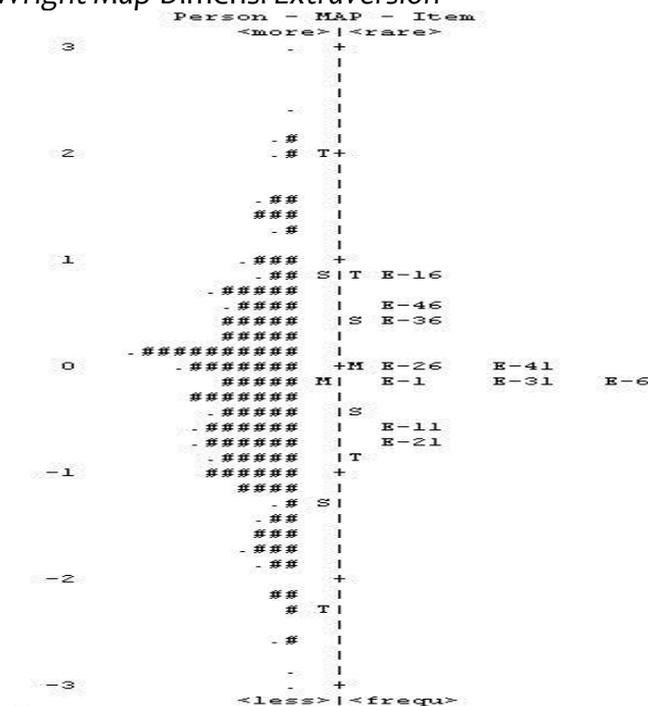


Tabel 3 Analisis Dimensi Extraversion

	Measure	S.E.Measure	Infit MS	Outfit MS	Point measure corr.
E16	0.80	0.07	1.24	1.32	0.48
E46	0.64	0.06	0.95	0.93	0.66
E36	0.39	0.06	1.07	1.05	0.67
E26	0.00	0.06	0.90	0.89	0.70
E41	-0.04	0.06	1.16	1.13	0.66
E6	-0.11	0.06	0.93	0.92	0.74
E1	-0.17	0.06	0.86	0.86	0.68
E31	-0.20	0.06	1.10	1.12	0.67
E11	-0.61	0.06	0.94	0.99	0.54
E21	-0.70	0.06	0.82	0.80	0.68

Pada analisis persebaran butir, didapati beberapa butir dengan nilai logit yang memiliki perbedaan nilai yang dekat seperti butir E-26=(0.00) dan E-41=(-0.04), serta E-1=(-0.17) dan E-31=(-0.20). Butir-butir dengan nilai logit yang berdekatan akan memberikan informasi abilitas orang yang juga hampir sama serta lebih lanjut akan kesusahan untuk benar-benar membedakan abilitas individu dengan lebih baik, hal ini serupa dengan butir-butir pada dimensi *emotional stability*, meskipun tidak terlalu banyak butir yang bernilai terlalu berdekatan.

Gambar 2 Wright Map Dimensi Extraversion



### Conscientiousness

Batasan *point measure correlation* yang disarankan bernilai lebih dari 0.3, (Linacre, 2011), sementara pada dimensi ini nilainya berkisar pada rentang 0.40-0.65 dan tidak ditemukan adanya nilai yang negatif. Hal ini menandakan bahwa butir-butir yang ada mampu dengan baik untuk membedakan orang-orang dengan abilitas yang rendah dan tinggi sehingga tidak menyesatkan. Apabila dilihat dari nilai logit, diketahui bahwa butir yang cenderung sulit untuk mendapatkan persetujuan oleh responden terutama dengan abilitas yang rendah adalah butir C-23 = 0.46, sementara itu butir C-38 = (-0.58) merupakan butir yang cenderung paling mudah mendapatkan persetujuan responden. Melalui nilai *infit-outfit meansquare*, menunjukkan jika butir-butir di dalam dimensi ini fit atau sesuai dengan model, dengan batas yang menurut (Bond & Fox, 2015) berkisar antara 0.5-1.5. Namun pada butir C-28 nilai *infit-outfit meansquare* menunjukkan nilai yang melebihi batas, hal ini dapat diasumsikan jika butir ini *underfit* dengan permodelan yang ideal dan hal ini bisa berdampak pada analisis dimensionalitas karena hubungannya dengan validitas konstruk. Setelah peneliti melakukan analisis ulang tanpa mengikutkan butir C-28, didapati nilai uji dimensionalitas yang meningkat dari 36.6% menjadi 39.9% . Sementara pada butir C-8 nilai *infit-outfit meansquare* yang mulai mendekati batasan atas ideal menandakan jika menurut rasch model butir ini tidak terlalu baik validitasnya untuk mengukur apa yang ingin diukur jika dibandingkan dengan butir yang semakin mendekati nilai 1 seperti butir C-23 atau C-33.

**Tabel 4**

Analisis Dimensi Conscientiousness

	<i>Measure</i>	<i>S.E.Measure</i>	<i>Infit MS</i>	<i>Outfit MS</i>	<i>Point measure corr.</i>
C23	0.46	0.06	1.04	1.04	0.59
C48	0.36	0.06	0.68	0.69	0.65
C28	0.35	0.06	1.52	1.60	0.40
C13	0.17	0.06	0.82	0.80	0.63
C43	0.06	0.06	0.80	0.80	0.65
C8	-0.02	0.06	1.41	1.44	0.51
C3	-0.21	0.07	0.70	0.68	0.64
C18	-0.29	0.07	1.00	1.15	0.52
C33	-0.29	0.07	0.81	0.82	0.61
C38	-0.58	0.07	1.25	1.22	0.49

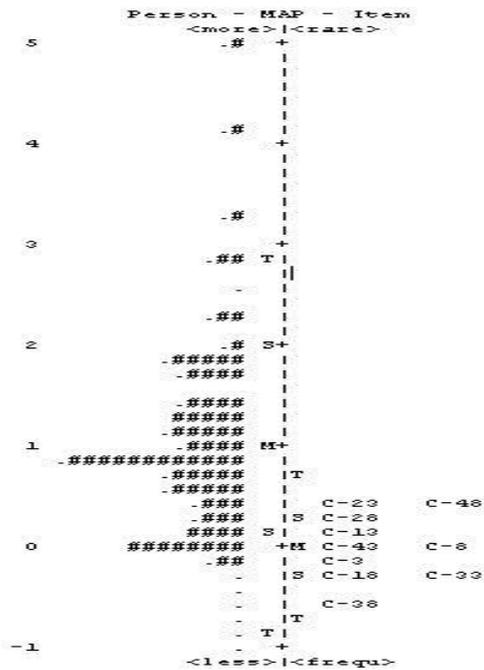
Pada analisis persebaran butir, didapati beberapa butir dengan nilai logit yang memiliki jarak nilai yang dekat seperti butir C-48=0.36 dan C-28=0.35, serta C-43=0.06 dan C-8=(-0.02). Butir-butir dengan nilai logit yang berdekatan dinilai akan memberikan informasi abilitas orang yang juga hampir sama sehingga tidak efisien jika dibandingkan dengan instrumen yang memiliki tingkat kesulitan butir yang bervariasi, lebih lanjut akan



kesusahan untuk benar-benar membedakan abilitas individu dengan lebih baik. Meskipun demikian ditemukan juga butir dengan nilai logit yang sama, C-18=(-0.29) dan C-33=(-0.29). Jika mengacu pada nilai indeks separasi *item-person*, dimensi ini memiliki nilai yang paling rendah, hal ini dapat terjadi dikarenakan butir-butir yang memiliki daya diskriminasi cenderung rendah dan tingkat kesulitan yang terlalu mudah jika dibandingkan dengan persebaran responden sehingga kurang dapat menjangkau individu-individu dengan abilitas *conscientiousness* yang tinggi.

**Gambar 3**

Wright Map Dimensi *Conscientiousness*



**Agreeableness**

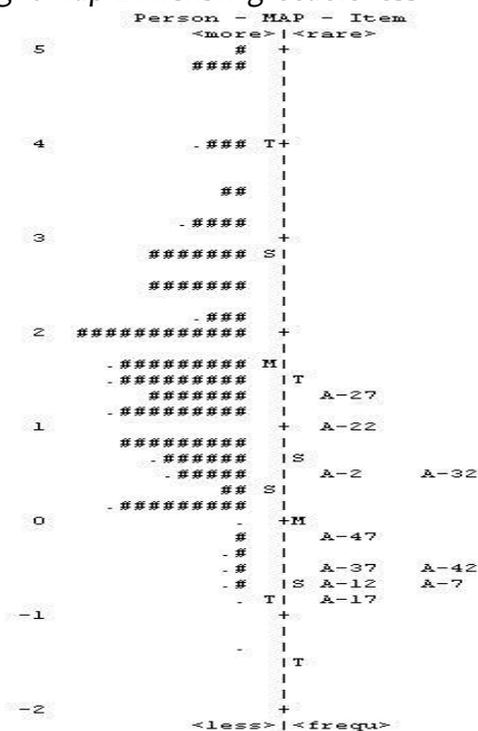
Batasan *point measure correlation* yang disarankan seharusnya bernilai lebih dari 0.3, jika mengacu pada (Linacre, 2011), sementara pada dimensi ini nilainya berkisar pada rentang 0.45-0.66 dan tidak ditemukan adanya nilai yang negatif. Hal ini mengindikasikan jika butir-butir yang ada mampu dengan baik untuk membedakan orang-orang dengan abilitas yang rendah dan tinggi sehingga membuat kebingungan. Apabila dilihat dari nilai logit, diketahui bahwa butir yang cenderung sulit untuk mendapatkan persetujuan oleh responden terutama dengan abilitas yang rendah adalah butir A-27 = 1.35, sementara pada butir A-17 = (-0.79) ditemukan merupakan butir yang cenderung paling mudah mendapatkan persetujuan responden. Melalui nilai *infit-outfit meansquare*, diindikasikan jika butir-butir di dalam dimensi ini fit atau sesuai dengan model, dengan batas yang menurut (Bond & Fox, 2015) berkisar antara 0.5-1.5

**Tabel 5**  
Analisis Dimensi Agreeableness

	Measure	S.E.Measure	Infit	Outfit	Point measure corr.
A27	1.35	0.07	1.30	1.34	0.47
A22	0.97	0.07	1.26	1.34	0.56
A32	0.47	0.07	1.29	1.33	0.58
A2	0.44	0.07	1.14	1.16	0.63
A47	-0.15	0.08	0.81	0.83	0.63
A37	-0.42	0.08	0.70	0.69	0.66
A42	-0.42	0.08	0.74	0.78	0.62
A7	-0.71	0.09	0.63	0.60	0.66
A12	-0.73	0.09	1.34	1.34	0.45
A17	-0.79	0.09	0.70	0.68	0.62

Dari persebaran pada *wright map*, diketahui jika persebaran butir telah cukup bervariasi sehingga lebih mampu untuk menjangkau dan mendiskriminasikan individu-individu dengan lebih efisien dibandingkan dimensi *emotional stability* dan juga *extraversion*, hal ini juga dapat ditinjau dari nilai indeks separasi meskipun dapat ditingkatkan lagi daya jangkau dan variasinya. Butir-butir seperti A-37=(-0.42) dan A-42=(-0.42) dinilai tidak cukup memiliki untuk memiliki perbedaan nilai yang variatif untuk lebih mudah membedakan individu-individu pada aspek *agreeableness*-nya.

**Gambar 4**  
Wright Map Dimensi Agreeableness





## Intellect

Apabila ditinjau dari *batas point measure correlation* yang disarankan bernilai lebih dari 0.3, (Linacre, 2011), maka butir-butir yang ada di dimensi ini mampu dengan baik untuk membedakan orang-orang dengan abilitas yang rendah dan tinggi sehingga tidak menyesatkan, dengan rentang nilai pada 0.43-0.74. Tidak adanya nilai negatif juga menandakan bahwa butir-butir yang ada mampu dengan baik untuk membedakan orang-orang dengan abilitas yang rendah dan tinggi sehingga tidak menyesatkan. Apabila dilihat dari nilai logit, diketahui bahwa butir yang cenderung sulit untuk mendapatkan persetujuan oleh responden terutama dengan abilitas yang rendah adalah butir I-40 = 1.27, sementara itu butir I-45 = (-0.58) merupakan butir yang cenderung paling mudah mendapatkan persetujuan responden. Melalui nilai *infit-outfit meansquare*, menandakan jika butir-butir di dalam dimensi ini fit atau sesuai dengan model, dengan batas yang menurut (Bond & Fox, 2015) berkisar antara 0.5-1.5. Jika diamati dari nilai logitnya, butir I-15 = (-0.45) dan I-30 = (-0.48) merupakan butir dengan nilai logit yang berdekatan. Hal ini mengindikasikan kesamaan tingkat kesulitan butir dan akan mempersulit untuk mendiskriminasi individu-individu dengan abilitas di dimensi *intellect* yang serupa. Nilai *infit-outfit* pada butir I-25 yang mendekati batasan bawah di satu sisi juga mengindikasikan jika butir ini cenderung overfit dengan model yang ada. Meskipun menurut (Bond & Fox, 2015) *overmisfit* dalam instrumen psikometri dapat lebih di toleransi jika dibandingkan dengan *underfit*, karena jika demikian maka butir tersebut tidak memenuhi validitas konstruk yang akan menyumbang pada analisis dimensionalitas yang buruk (kurang dari 40%).

**Tabel 6**

Analisis Dimensi *Intellect*

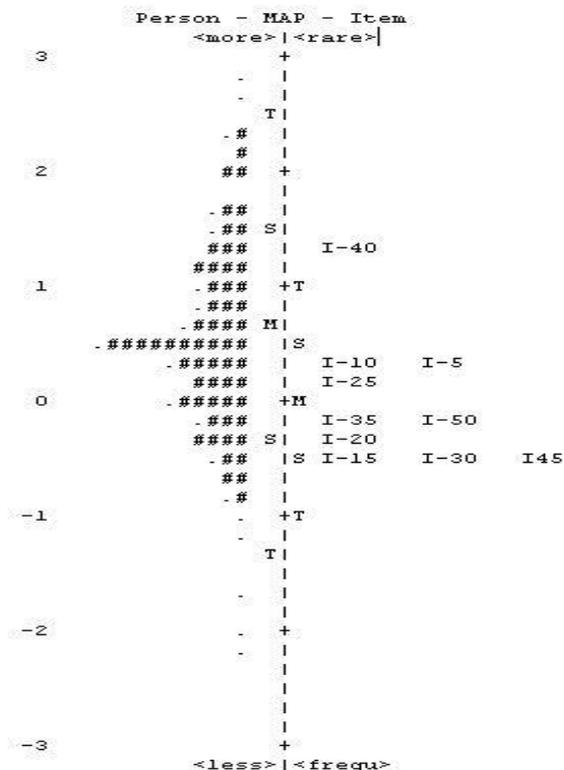
	<i>Measure</i>	<i>S.E.Measure</i>	<i>Infit</i>	<i>Outfit</i>	<i>Point measure corr.</i>
I40	1.27	0.07	1.19	1.17	0.52
I10	0.39	0.07	1.22	1.25	0.50
I5	0.28	0.07	0.87	0.89	0.56
I25	0.21	0.07	0.53	0.54	0.72
I50	-0.10	0.07	0.69	0.70	0.74
I35	-0.19	0.07	0.84	0.84	0.54
I20	-0.36	0.07	1.21	1.20	0.53
I15	-0.45	0.07	0.85	0.82	0.69
I30	-0.48	0.07	1.28	1.26	0.54
I45	-0.58	0.07	1.37	1.41	0.43

Ditinjau dari persebaran butir di *wright map*, variasi kesulitan butir dinilai telah cukup memadai individu-individu dengan variasi abilitas dari *intellect* yang beragam, meskipun dapat ditingkatkan lagi. Butir-butir pada dimensi ini dinilai lebih memiliki variasi

yang lebih baik jika dibandingkan dimensi *emotional stability* atau *extraversion* sebelumnya.

Gambar 5

Wright Map Dimensi Intellect



### DIFF (DIFFerential Item Functioning)

Butir-butir instrumen bisa saja terindikasi oleh bias, dalam analisis ini peneliti mencari butir yang berpotensi untuk bias berdasarkan jenis kelamin, peneliti memilih dua variabel dalam jenis kelamin yang umum digunakan pada budaya di Indonesia, laki-laki dan perempuan. Butir yang dinyatakan terindikasi bias dapat dilihat berdasarkan nilai mantelnya, dengan titik kritis 3.84 (Penfield, 2013).

Pada dimensi *agreeableness* terdapat 1 butir yang terdeteksi bias jenis kelamin, A-32 = 4.91. Berikutnya pada dimensi *conscientiousness*, ditemukan butir yang memiliki nilai mantel yang menunjukkan adanya bias, C-38 = 5.63. 2 butir pada dimensi *extraversion* yang memiliki nilai mantel lebih dari 3.84, yaitu E-6 = 7.36 dan E41= 8.97. Dimensi *intellect*, memiliki 1 butir yang teridentifikasi bias, butir I-30 = 5.47. Pada dimensi *emotional stability*, berikutnya butir EM-44 = 4.67, EM-14 = 5.51, EM-29 = 4.53 dan EM-9 = 10.42 merupakan butir yang teridentifikasi jenis kelamin dengan EM-9 dengan butir yang memiliki nilai paling tinggi.

### Pembahasan

Butir-butir pada kelima dimensi secara garis besar menunjukkan nilai *point measure correlation* yang memuaskan dengan nilai diatas 0.3 dan tidak ada satu butir pun memiliki



nilai negatif. Pada pendekatan CTT *point measure correlation* identik keberfungsannya dengan *item discrimination* indeks, hanya saja menurut (Smiley, 2015) analisis dari *rasch model* bertujuan untuk memberikan pengukuran yang lebih adil dengan pengukuran pada level interval yang mana tidak diakomodai melalui pendekatan CTT. Nilai *outfit-infit* sendiri merupakan nilai residual antara model dengan data empirik yang ada, (Bond & Fox, 2015), semakin besar diskrepansi residual yang ada maka tentunya berpengaruh pada validitas konstruksinya.

Apabila nilai *infit-outfit* terlalu fit dengan model (*overmisfit*) maka nilainya akan semakin mendekati 0, sebaliknya jika kurang fit dengan model (*underfit*) maka nilainya semakin menjauh dari 1. Seluruh butir dimulai dari dimensi *emotional stability* hingga *intellect* yang dianalisis, memiliki ketepatan dengan ambang nilai standar yang disarankan, yang mengindikasikan butir-butir instrumen ini pada setiap dimensinya telah memiliki pernyataan-pernyataan yang telah sesuai dengan konstruksi teori yang ada meskipun pada butir C-28 ditemukan nilai *infit-outfit*nya yang diatas 1.5, apabila peneliti gugurkan, analisis pada dimensionalitas meningkat sebesar 3.33% menjadi 39.9%, namun apabila mengacu pada penelitian sebelumnya, sebenarnya peneliti tidak melihat sebuah hal yang besar, mengingat butir C-28 memiliki informasi psikometri yang baik pada saat melalui proses adaptasi pada studi milik (Akhtar & Azwar, 2019). Secara garis besar, jika melihat hasil uji *infit-outfit meansquare*, ke empat dimensi lainnya yaitu, *emotional stability*, *extraversion*, *intellect*, hingga *agreeableness* telah mengindikasikan dimensionalitas yang baik (unidimensional) dengan batasan diatas 40%, meskipun nilainya berkisar antara 41% - 46.2% yang menandakan tidak berada terlalu jauh dari batas 40%. Namun sebenarnya batasan minimal dalam asumsi unidimensionalitas berada pada batasan 20% seperti halnya studi uji dimensionalitas milik (Misbach & Sumintono, 2011; Zahirah & Susanto, 2021), sementara pada studi lainnya digunakan asumsi batasan 40%, seperti pada milik (Natanael, 2021). Bagaimanapun juga apabila peneliti mengacu pada (Linacre, 2011) dan meninjau nilai *unexplained variance in 1st contrast* pada keseluruhan dimensi, didapatkan nilai lebih dari 2, yang mana hal ini tidak ditemukan pada *unexplained variance in 2nd contrast* hingga *5th contrast* di keseluruhan dimensi, hal ini menunjukkan bukti tersendiri jika kelima dimensi yang diuji secara terpisah telah memenuhi syarat unidimensionalitasnya masing-masing.

Persebaran *person-item* pada *wright map* menunjukkan dimensi *emotional stability*, *extraversion*, *conscientiousness*, *agreeableness* dan *intellect* memiliki variasi butir yang tidak terlalu variatif dikarenakan terdapat beberapa butir dengan tingkat abilitas yang serupa, bahkan pada beberapa dimensi ditemukan butir dengan tingkat abilitas sama. Hal ini dapat ditunjukkan dengan nilai indeks separasi *person* yang berada dibawah 3. Sorotan lebih lanjut sebenarnya ditujukan kepada dimensi *emotional stability*, *extraversion*, dan *conscientiousness* yang juga memiliki indeks separasi *item* tidak begitu memuaskan (dibawah 3), hal ini mengindikasikan jika sampel yang diperoleh sebenarnya tidak cukup beragam untuk menggambarkan variasi tingkat kesulitan butir yang ada, sementara ketidakmerataan sampel terjadi dengan indikasi indeks separasi butir pada dimensi *agreeableness* = 8.58 dan *intellect* = 7.42 yang sangat memuaskan, dengan demikian

---

responden-responden yang didapat pada penelitian ini merepresentasikan orang-orang dengan *agreeableness* dan *intellect* yang lebih beragam jika dibandingkan dengan dimensi lainnya. Terkait dengan analisis reliabilitas *person-items* kelima dimensi yang ada memiliki nilai reliabilitas person-item yang memuaskan, menurut (Linacre, 2011) hal ini menandakan keajegan dari responden dan butir itu sendiri. Menurut (Schumacker & Smith, 2007) dengan penggunaan standar eror yang bersifat individu, analisis reliabilitas *rasch model* memiliki keunggulan pada kejelasan diagnosis tersendiri jika dibandingkan CTT yang memiliki satu standar eror untuk semua, selain daripada itu perbedaan lainnya yang juga dapat diamati dari analisis reliabilitas terdapat.

Kelima dimensi pada alat ukur BFM-50 terindikasi memiliki bias jenis kelamin dengan proporsi kuantitas butir yang berbeda-beda. Dimensi yang paling banyak mengandung bias adalah dimensi *emotional stability* sebanyak empat butir (EM-44, EM-14, EM-29 dan EM-9), kemungkinan hal ini dapat terjadi karena adanya bias terkait persepsi kolektif pada masyarakat Indonesia terhadap pemilihan diksi hasil terjemahan dari versi asli (bahasa Inggris). Sebagai sebuah gambaran pada butir EM-9 yang telah melalui uji adaptasi pada penelitian (Akhtar & Azwar, 2019), berbunyi "merasa tenang hampir setiap saat", dengan adanya bias dapat kita simpulkan jika butir ini memiliki tingkat kesulitan / kemudahan yang berbeda jika dikerjakan oleh perempuan dan laki-laki, yang semestinya tingkat kesulitan tersebut tidak seharusnya ditentukan oleh jenis kelamin melainkan pada substansi kalimat dalam butirnya agar diperoleh pengukuran yang setara. Tentunya memoDIFikasi bunyi butir dengan pemilihan diksi yang berbeda perlu dipertimbangkan untuk dilakukan untuk didapatkan butir-butir yang bebas dari bias jenis kelamin.

Dari penelitian yang sudah dilakukan masih terdapat keterbatasan yaitu analisis data yang dilakukan dengan cara terpisah pada setiap dimensinya. Menurut (Baghaei, 2012), menganalisis data menggunakan asumsi *rasch model* dengan terpisah pada setiap dimensinya dapat mengakibatkan pengukuran yang kurang akurat jika dibandingkan analisis yang serentak pada semua dimensinya. Selain daripada itu teknik dan jumlah sampling sebenarnya dapat ditingkatkan lagi, mengingat terdapat persebaran individu-individu yang tidak merata dengan indikasi nilai indeks separasi butir yang rendah dan tinggi pada beberapa dimensi, peneliti juga menyarankan untuk melakukan seleksi responden dengan lebih selektif, dalam hal ini adalah proses pengguguran responden yang tidak sesuai dengan analisis *infit-outfit*, bahkan yang nilainya mendekati 1.5 atau 0.5, hal ini dilakukan untuk menambah presisi pengukuran dan mengurangi pengotor yang dapat mengganggu analisis validitas.

Berikutnya untuk ide pengembangan hasil temuan ini, peneliti berpikir untuk adanya perbaikan jika dirasa memang perlu, pada butir-butir yang terindikasi bias jenis kelamin dan butir-butir yang kurang memiliki variasi tingkat kesulitan yang beragam untuk melalui proses peninjauan ulang. Diharapkan juga pada penelitian-penelitian berikutnya dalam proses adaptasi maupun validasi alat ukur, dapat dipertimbangkan untuk melakukan analisis statistik yang kolaboratif antara pengukuran berbasis CTT dengan IRT dalam hal ini *rasch model* dikarenakan, kedua pendekatan ini masing-masing akan



memberikan informasi psikometri khususnya tersendiri, yang akan semakin mempermudah peneliti untuk merancang atau mengkonstruksi instrumen dengan lebih percaya diri dikarenakan informasi yang terbilang lebih banyak antara kedua pendekatan untuk saling dipertimbangkan dan melengkapi satu dengan lainnya.

Apabila menjumpai instrumen yang bersifat multidimensional peneliti berharap untuk dipertimbangkan dalam penggunaan aplikasi analisis *rasch model* yang mengakomodasi permodelan statistik untuk multidimensional *rasch model*. Pada pengaplikasiannya ditemukan beberapa peneliti menggunakan aplikasi ConQuest yang dikembangkan oleh (Adams, Wu, Cloney, & Wilson, 2020) untuk menganalisis data bermodelkan *rasch* yang multidimensional, seperti pada penelitian milik (Yan & Mok, 2012) yang memvalidasi instrumen *Coping Scale for Chinese Athletes (CSCA)* dengan komposisi empat dimensi, tentunya hal ini menarik untuk menjadi pertimbangan penelitian berikutnya pada alat ukur kepribadian. Berkaitan dengan *Differential item functioning* tentunya tidak terbatas hanya pada jenis kelamin (laki-laki dan perempuan) namun dapat juga digunakan variabel lainnya yang setara, hal ini dapat menjadikan DIF sebagai analisis alternatif lain untuk mendeteksi bias dalam sebuah instrumen psikometri.

## Kesimpulan

Berdasarkan proses validasi IPIP-BFM 50 menggunakan *rasch model* secara garis besar dapat disimpulkan bahwa pada dimensi *emotional stability*, *extraversion*, *conscientiousness*, *agreeableness*, dan *intellect* memiliki butir-butir yang telah memenuhi syarat untuk setidaknya bersesuaian mengikuti model *rasch* yang ada. Harga *outfit-infit meansquare* untuk ke-49 butir berkisar antara 0,5 - 1,5, dengan demikian residu yang ada pun tidak terlalu berpengaruh terhadap akurasi butir-butir pada dimensi-dimensi tersebut dengan pemodelan *rasch*. Butir C-28 yang mengalami *underfit* dengan nilai *infit-outfit meansquare* yang sedikit melebihi diatas 1.5 mengindikasikan jika butir ini dapat dikategorikan sebagai pengotor yang bisa merusak validitas konstruk dari dimensi *conscientiousness*, meskipun pada proses validasi studi sebelumnya butir C-28 tidak menunjukkan adanya indikasi bermasalah. Terdapat dimensi *emotional stability*, *extraversion*, dan *conscientiousness*, *agreeableness*, dan *intellect* yang memiliki butir-butir dengan tingkat kesulitan yang kurang variatif atau beragam bahkan dengan nilai logit sama. Tentunya hal ini berpengaruh efisiensi pada daya beda instrumen ini yang kurang baik untuk membedakan orang-orang dalam sebuah kelompok dengan abilitas yang beragam. Seluruh dimensi pada alat ukur ini memiliki setidaknya satu butir dengan bias jenis kelamin, dengan *emotional stability* sebagai dimensi dengan jumlah butir paling banyak (EM-44, EM-14, EM-29, dan EM-9).

Peneliti menyarankan adanya perubahan butir yang terindikasi bias dan untuk memperbaiki adanya variasi tingkat kesulitan butir yang tidak terlalu baik. Selain daripada itu, pada penelitian-penelitian serupa berikutnya, diharapkan dapat mempertimbangkan untuk tidak ragu menggunakan pendekatan kolaboratif antara paradigma CTT dan *rasch model* seperti pada studi adaptasi, pengembangan instrumen, validasi, dkk. Penggunaan

analisis *rasch model* secara simultan apabila bertemu dengan instrumen yang multidimensi tentu juga menjadi catatan dalam penelitian ini dikarenakan abilitasnya yang lebih baik jika dibandingkan dengan analisis secara terpisah. Berkaitan dengan *DIFferential item functioning* tentunya tidak terbatas hanya pada jenis kelamin (laki-laki dan perempuan) namun dapat juga dipertimbangkan variabel lainnya yang memungkinkan untuk dilakukan uji DIF.

## Referensi

- Adams, R. J., Wu, M. L., Cloney, D., & Wilson, M. R. (2020). *ACER ConQuest: Generalised Item Response Modelling Software [Computer software] Version 5*. Victoria: Australian Council for Educational Research.
- Akhtar, H., & Azwar, S. (2019). Indonesian Adaptation and Psychometric Properties Evaluation of the Big Five Personality Inventory: IPIP-BFM-50. *Jurnal Psikologi*, 46(1), 32. <https://doi.org/10.22146/jpsi.33571>
- Ashley, L., Smith, A. B., Keding, A., Jones, H., Velikova, G., & Wright, P. (2013). Psychometric evaluation of the Revised Illness Perception Questionnaire (IPQ-R) in cancer patients: Confirmatory factor analysis and Rasch analysis. *Journal of Psychosomatic Research*, 75(6), 556–562. <https://doi.org/10.1016/j.jpsychores.2013.08.005>
- Babarović, T., & Šverko, I. (2013). HEXACO dimenzije ličnosti u hrvatskom uzorku. *Drustvena Istrazivanja*, 22(3), 397–411. <https://doi.org/10.5559/di.22.3.01>
- Baghaei, P. (2012). The application of multidimensional rasch models in large scale assessment and validation: An empirical example. *Electronic Journal of Research in Educational Psychology*, 10(1), 233–252. <https://doi.org/10.25115/ejrep.v10i26.1493>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model* (Routledge (ed.); III).
- Brandt, S., Moulton, M., & Duckor, B. (2015). Advances in Rasch modeling: New applications and directions Guest Editorial. *Psychological Test and Assessment Modelling*, 57(3), 338–341.
- Costa, P. (2018). *Neo PI-R professional manual*. January 1992.
- Cupani, M., & Lorenzo-Seva, U. (2016). The development of an alternative IPIP inventory measuring the Big-Five factor markers in an Argentine sample. *Personality and Individual Differences*, 91(March), 40–46. <https://doi.org/10.1016/j.paid.2015.11.051>
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2), 192–203. <https://doi.org/10.1037/1040-3590.18.2.192>
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's "IPIP" Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39(2), 317–329. <https://doi.org/10.1016/j.paid.2005.01.011>



- Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, 46(3), 380–393. <https://doi.org/10.1016/j.ijnurstu.2008.10.007>
- Ingram, P. B., Boan-Lenzo, C., & Vuyk, M. A. (2013). Openness/Intellect in a 50-Item IPIP Instrument. *Journal of Psychoeducational Assessment*, 31(6), 600–605. <https://doi.org/10.1177/0734282913481448>
- Jabrayilov, R., Emons, W. H. M., & Sijtsma, K. (2016). Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Applied Psychological Measurement*, 40(8), 559–572. <https://doi.org/10.1177/0146621616664046>
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78–89. <https://doi.org/10.1016/j.jrp.2014.05.003>
- Lim, B. C., & Ployhart, R. E. (2006). Assessing the convergent and discriminant validity of Goldberg's international personality item pool: A multitrait-multimethod examination. *Organizational Research Methods*, 9(1), 29–54. <https://doi.org/10.1177/1094428105283193>
- Lin, C. Y., Hwang, J. S., Wang, W. C., Lai, W. W., Su, W. C., Wu, T. Y., Yao, G., & Wang, J. Der. (2019). Psychometric evaluation of the WHOQOL-BREF, Taiwan version, across five kinds of Taiwanese cancer survivors: Rasch analysis and confirmatory factor analysis. *Journal of the Formosan Medical Association*, 118(1P2), 215–222. <https://doi.org/10.1016/j.jfma.2018.03.018>
- Linacre, J. M. (2011). *Winsteps Help for Rasch Analysis*. <http://homes.jcu.edu.au/~edtgb/%5Cnpapers3://publication/uuid/D56B724A-62FF-4D00-84E1-ECC888298B70>
- Linacre, J. M. (2022). *Winsteps® Rasch measurement computer program (Version 5.2.2)*. Diambil kembali dari <https://www.winsteps.com/>
- Mallinckrodt, B., Miles, J. R., & Recabarren, D. A. (2016). Using Focus Groups and Rasch Item Response Theory to Improve Instrument Development. In *Counseling Psychologist* (Vol. 44, Issue 2). <https://doi.org/10.1177/0011000015596437>
- Maples, J. L., Guan, L., Carter, N. T., & Miller, J. D. (2014). A test of the international personality item pool representation of the revised NEO personality inventory and development of a 120-item IPIP-based measure of the five-factor model. *Psychological Assessment*, 26(4), 1070–1084. <https://doi.org/10.1037/pas0000004>
- Misbach, I. H., & Sumintono, B. (2011). PENGEMBANGAN DAN VALIDASI INSTRUME “PERSEPSI SISWA TEHADAP KARAKTER MORAL GURU” DI INDONESIA DENGAN MODEL RASCH. 59–66.
- Natanael, Y. (2021). Analisis Rasch model Indonesia Problematic Internet Use Scale (IPIUS). *Persona: Jurnal Psikologi Indonesia*, 10(1), 167–186. <https://doi.org/10.30996/persona.v10i1.4827>
- Oliveira, J. P. (2019). Psychometric Properties of the Portuguese Version of the Mini-IPIP five-Factor Model Personality Scale. *Current Psychology*, 38(2), 432–439. <https://doi.org/10.1007/s12144-017-9625-5>

- Penfield, R. D. (2013). *DIFas 5.0*.
- Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2012). *Breaking Free From The Limitation of Classical Test Theory: Developing and Measuring Information Systems Scales Using Item Response Theory*. December, 0–5.
- Sakib, N., Bhuiyan, A. K. M. I., Hossain, S., Al Mamun, F., Hosen, I., Abdullah, A. H., Sarker, M. A., Mohiuddin, M. S., Rayhan, I., Hossain, M., Sikder, M. T., Gozal, D., Muhit, M., Islam, S. M. S., Griffiths, M. D., Pakpour, A. H., & Mamun, M. A. (2020). Psychometric Validation of the Bangla Fear of COVID-19 Scale: Confirmatory Factor Analysis and Rasch Analysis. *International Journal of Mental Health and Addiction*. <https://doi.org/10.1007/s11469-020-00289-x>
- Sari, U. P. P. (2016). Relationship Between personality (Openness, Conscientiousness, Extraversion, Agreeableness, Emotional Stability) With Student's Environmental Moral Behavior. *Indonesian Journal of Environmental Education and Management*, 1(2), 1–13.
- Schultz, D. P., & Schultz, S. E. (2009). Theories of Personality. In *Wadsworth Cengage Learning*. Wadsworth Cengage Learning. <https://doi.org/10.1093/oxfordjournals.jbchem.a126528>
- Schumacker, R. E., & Smith, E. V. (2007). A rasch perspective. *Educational and Psychological Measurement*, 67(3), 394–409. <https://doi.org/10.1177/0013164406294776>
- Smiley, J. (2015). *Classical test theory or Rasch: A personal account from a novice user*. 19(1).
- Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, 8, 1–11. <https://doi.org/10.1186/1471-2288-8-33>
- Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch Measurement for the Development and Use of Affective Assessments in Science Education Research. *Science Education*, 98(4), 581–613. <https://doi.org/10.1002/sce.21118>
- Strus, W., Ciecuch, J., & Rowiński, T. (2014). The polish adaptation of the IPIP-BFM-50 questionnaire for measuring five personality traits in the lexical approach. *Roczniki Psychologiczne*, 17(2), 327–366.
- Sumintono, B. (2014). Model Rasch untuk Penelitian Sosial Kuantitatif. *Makalah Kuliah Umum Di Jurusan Statistika, ITS Surabaya, 21 November 2014, November 201*, 1–9. <http://deceng3.wordpress.com>
- Taasoobshirazi, G., Bailey, M. L., & Farley, J. (2015). Physics Metacognition Inventory Part II: Confirmatory factor analysis and Rasch analysis. *International Journal of Science Education*, 37(17), 2769–2786. <https://doi.org/10.1080/09500693.2015.1104425>
- Weiten, W. (2013). *Psychology Themes and Variations 10th Edition*.
- Yan, Z., Brubacher, S., Boud, D., & Powell, M. (2020). Psychometric properties of the Self-assessment Practice Scale for professional training contexts: evidence from confirmatory factor analysis and Rasch analysis. *International Journal of Training and Development*, 24(4), 357–373. <https://doi.org/10.1111/ijtd.12201>
- Yan, Z., & Mok, M. M. C. (2012). Validating the coping scale for Chinese athletes using



multidimensional Rasch analysis. *Psychology of Sport and Exercise*, 13(3), 271–279.  
<https://doi.org/10.1016/j.psychsport.2011.11.013>

Zahirah, A., & Susanto, H. (2021). Aplikasi model Rasch pada adaptasi skala personal fable remaja di Jawa Barat. *Persona: Jurnal Psikologi Indonesia*, 10(1), 63–80.  
<https://doi.org/10.30996/persona.v10i1.5097>

Zheng, L., Goldberg, L. R., Zheng, Y., Zhao, Y., Tang, Y., & Liu, L. (2008). Reliability and concurrent validation of the IPIP Big-Five factor markers in China: Consistencies in factor structure between Internet-obtained heterosexual and homosexual samples. *Personality and Individual Differences*, 45(7), 649–654.  
<https://doi.org/10.1016/j.paid.2008.07.009>