

Uji Validitas dan Reliabilitas *Multicultural Attitude Scale Questionnaire* (MASQUE) Versi Indonesia dengan Model Rasch

Livia Tanujaya, Ananta Yudiarto

Fakultas Psikologi, Universitas Surabaya, Surabaya, Indonesia

e-mail: liviatanujaya@gmail.com, ananta@staff.ubaya.ac.id

Abstract

Indonesia is a multicultural country that refers to the diversity of tribes, races, religions and beliefs. This study aims to enrich the development findings by providing psychometric information on the Munroe Multicultural Attitude Scale Questionnaire (MASQUE) using the Model Rasch perspective. This also introduces the application of the Model Rasch in analyzing the validity of the Munroe Multicultural Attitude Scale Questionnaire (MASQUE). This instrument has three aspects, namely (1) Knowledge, (2) Concern, and (3) Action and has a total of 18 items. Respondents used in this study were 210 students who were taken using simple random sampling method. The results of data analysis using the Model Rasch show that the Indonesian version of the MASQUE instrument has items that are fit and have quite a variety of item difficulty levels, but adjustments need to be made in order to facilitate a more varied personality and items that are spread out in more appropriate aspects.

Keywords: *Multiculturalism Attitude, Multiculturalism, MASQUE, Model Rasch*

Abstrak

Indonesia merupakan negara multikultural yang merujuk terhadap keragaman suku, ras, agama, dan kepercayaan. Penelitian ini bertujuan untuk memperkaya temuan pengembangan dengan memberikan informasi psikometri Munroe Multicultural Attitude Scale Questionnaire (MASQUE) menggunakan perspektif Model Rasch. Hal ini sekaligus dapat memperkenalkan penerapan Model Rasch dalam menganalisa validitas alat ukur Munroe Multicultural Attitude Scale Questionnaire (MASQUE). Instrumen ini memiliki tiga aspek, yaitu (1) Pengetahuan, (2) Kepedulian, dan (3) Tindakan serta memiliki total item sebanyak 18 butir. Responden yang digunakan pada penelitian ini adalah sebanyak 210 orang mahasiswa yang dipilih dengan menggunakan *purposive sampling*. Hasil analisis data dengan Model Rasch menunjukkan bahwa instrument MASQUE versi Bahasa Indonesia memiliki butir-butir yang sudah fit dan cukup bervariasi tingkat kesulitan itemnya, namun perlu dilakukan penyesuaian agar dapat memfasilitasi keragaman karakteristik orang yang lebih bervariasi dan item yang tersebar dalam aspek lebih sesuai.

Kata kunci: Sikap Multikulturalisme, Multikulturalisme, MASQUE, Model Rasch

I. Pendahuluan

Sebagai negara dengan penduduk terbanyak ke-empat di dunia dan memiliki asal usul sejarah yang kaya, Indonesia merupakan salah satu contoh negara multikultural. Multikultural tidak hanya terbatas pada suku bangsa dan etnisitas, namun juga pada latar belakang agama, jenis kelamin dan preferensi seksual (Zarbaliyev, 2017). Menurut Banks (1996, dalam Gabeyehu, 2022), multikulturalisme merupakan sebuah proses transformatif melalui sistem pendidikan yang dirancang khusus. Sistem pendidikan multikultural dijelaskan Banks sebagai penekanan pada pentingnya keragaman yang diyakini mempengaruhi perluasan dan kesetaraan pengalaman subjektif individu dalam hal pendidikan, hubungan sosial, hingga pembentukan identitas diri.

Indonesia bisa saja memiliki tatanan masyarakat multikultur, namun tidak serta merta menjadikan Indonesia negara yang menganut multikulturalisme (Parker, 2017). Sikap multikulturalisme dapat dilihat dari *prejudice* atau *judgement* serta kecenderungan tindakan yang diambil kelompok masyarakat bila berhadapan dengan kelompok lainnya yang berbeda. Sikap multikulturalisme yang terus ditingkatkan akan menurunkan *prejudice* maupun *judgement* dari kelompok masyarakat (Cobb et al, 2020). Namun, penilaian seseorang bukan hanya dibentuk dalam keluarga ataupun orang terdekat, akan tetapi adanya stereotip yang menjadi konstruksi sosial sehingga menjadi hambatan dari terciptanya kondisi multikulturalisme.

Untuk menjelaskan sikap multikulturalisme secara terukur, maka Munroe menciptakan sebuah alat ukur yaitu Munroe Multicultural Attitude Scale Questionnaire (MASQUE). Dasar teori dari MASQUE adalah teori pendekatan transformasional Banks dan taksonomi Bloom (Munroe, 2006 dalam Permatasari et al., 2020). Awalnya, alat ukur ini diciptakan sebagai pendukung pendidikan multikultural, karena teori Banks dan konsep taksonomi Bloom sepakat bahwa perubahan sikap dan perilaku dapat dipahami melalui tiga aspek, yaitu *knowledge*, *care*, dan *act* yang kemudian juga menjadi tiga dimensi yang diperkenalkan pada MASQUE. Alat ukur ini tidak menyertakan kewarganegaraan tertentu sehingga tetap dapat digunakan secara luas.

Saat pertama kali MASQUE diciptakan pada tahun 2006, sudah ada alat ukur lain hanya saja digunakan lebih spesifik pada guru, konselor dan sekolah (Permatasari et al, 2020). Alat ukur yang mengukur perilaku dan sikap multikulturalisme secara bersamaan sangat sedikit. Selain itu MES, MAKSS, dan MCKAS hanya mengukur pada kelompok masyarakat yang sangat spesifik sehingga penyesuaian yang perlu dilakukan akan cukup banyak.

Kebutuhan untuk mengadaptasi MASQUE menjadi penting karena tiap negara memiliki pemaknaan yang berbeda dalam melihat multikulturalisme (Hoon, 2017). Permatasari et al (2020) melakukan sebuah penelitian untuk menciptakan MASQUE Versi Bahasa Indonesia. Hal ini dikarenakan MASQUE tidak bebas budaya sehingga perlu dialihbahasakan agar dapat dipahami juga oleh masyarakat Indonesia yang mungkin belum dapat berbahasa Inggris. Setelah melalui proses transadaptasi, alat ukur ini juga melalui proses uji validitas dengan menggunakan *Confirmatory Factor Analysis* (CFA). Hasilnya, beberapa item dalam alat ukur perlu dieliminasi karena validitas konstruksinya kurang baik.

Penelitian ini akan melakukan uji validitas alat ukur MASQUE dengan analisis metode Rasch untuk memberikan kontribusi temuan mengenai alat ukur ini sehingga

MASQUE Versi Indonesia dapat digunakan secara lebih meluas pada masyarakat Indonesia. Sejauh ini, validasi menggunakan Model Rasch belum pernah dilakukan untuk MASQUE adaptasi Bahasa Indonesia, dan belum banyak digunakan dalam penelitian validitas secara umum.

Model Rasch merupakan salah satu model pengukuran psikologis yang menyediakan kerangka matematika yang digunakan untuk membandingkan data dari hasil tes (Azizah & Wahyuningsih, 2020). Analisis Model Rasch diterapkan untuk meningkatkan ketepatan konstruk instrumen, mengetahui kualitas instrumen, dan menghitung kemampuan responden. Model Rasch menggunakan pendekatan *Item Response Theory* (IRT), yang menganalisis interaksi antara responden dengan items. IRT tidak bergantung pada item tertentu dengan kemampuan orang saat mengerjakan item tersebut. IRT memiliki tiga parameter logistik, yang salah satunya merupakan Model Rasch (Pratama, 2020).

Model Rasch menggunakan logit atau log probabilitas responden dalam menjawab item. Hal ini dapat membantu memperkirakan tingkat kesalahan yang berkaitan dengan item. Dalam Model Rasch, skala *rating scale* diposisikan sebagai skala ordinal. Salah satu asumsi dasarnya bahwa individu dengan kemampuan yang tinggi memiliki probabilitas besar untuk menjawab benar dibandingkan individu lainnya.

Model Rasch mampu menjadi jawaban untuk kelemahan berbagai metode analisis yang saat ini sudah banyak dilaksanakan. Salah satunya adalah *Classical Test Theory* (CTT) yang umum dilakukan pada penelitian ilmu sosial, termasuk psikologi, contohnya pada penelitian yang dilakukan oleh Setiawati et al (2021). CTT memiliki kelemahan dalam pengukuran, yaitu saat tes yang berbeda digunakan untuk mengukur hal yang sama dan dilakukan pada siswa yang berbeda, maka perbandingan antara item dengan responden tidak valid.

Dengan pertimbangan tersebut, peneliti menilai penting untuk melakukan validasi alat ukur MASQUE versi Indonesia dengan menggunakan Model Rasch. Hal ini karena setiap alat ukur memiliki perbedaan respon dari masing-masing responden, maka penelitian uji validitas diperlukan dengan menyesuaikan respon dari setiap negara dan bahasa. Model Rasch merupakan jenis analisis yang dapat menguji validitas dan reliabilitas instrumen alat ukur dengan menguji *person* dan *item* secara simultan. Model Rasch juga memiliki beberapa kelebihan karena sesuai dengan prinsip pengukuran yaitu mampu memberikan skala linier dengan interval yang sama, dapat melakukan prediksi terhadap data yang hilang, serta mampu mendeteksi ketidaktepatan model.

II. Metode Penelitian

Kriteria inklusi partisipan yang dapat terlibat dalam penelitian ini adalah laki-laki dan perempuan berusia mulai 18 tahun serta merupakan mahasiswa di Indonesia. Jumlah responden nantinya akan mengikuti aturan *sample size* dalam Model Rasch yakni kurang lebih 150-250 responden, dengan logit $\frac{1}{2}$ atau 95% atau *high stake*. Sedangkan kriteria eksklusi partisipan yakni dewasa berusia 18 tahun ke atas, laki-laki maupun perempuan yang bukan merupakan mahasiswa.

Instrumen yang digunakan dalam penelitian ini merupakan *Multicultural Attitude Scale Questionnaire* (MASQUE) oleh Munroe & Pearson (2006) yang diadaptasi dengan menggunakan Bahasa Indonesia oleh Permatasari et al. (2020) yang juga memberikan izin kepada penulis untuk melakukan pengukuran validitas dan reliabilitas menggunakan Model Rasch.. Instrumen ini merupakan instrumen multidimensional berdasarkan tiga dimensi yaitu: Know (Pengetahuan), Care (Kepedulian), dan Act (Tindakan) yang pilihan jawabannya menggunakan Skala Likert dengan rentang skor 1 (Sangat Tidak Setuju) hingga 6 (Sangat Setuju) . Permatasari et al. (2020) memodifikasi menjadi 14 item dari 18 item dengan mempertimbangkan validitas konstruk sebelum dan sesudah serta reliabilitas sebelumnya yang kurang baik pada dimensi Act dan reliabilitas keseluruhan setelahnya ($\alpha = 0,83$).

Tabel I. Item Terjemahan (Permatasari et al., 2020)

No.	Butir
Dimensi Know (Pengetahuan)	
1.	Saya menyadari bahwa rasisme terjadi di sekitar saya.
2.	Saya menyadari adanya batasan sosial.
3.	Saya memahami bahwa kepercayaan agama bisa berbeda-beda.
4.	Saya memahami bahwa kecenderungan seksual bisa berbeda-beda.
5.	Saya memahami adanya ketidakadilan terhadap perempuan.
6.	Saya menerima kenyataan bahwa terdapat bahasa selain bahasa Indonesia yang digunakan.
Dimensi Kepedulian (Care)	
7.	Saya menganggap penting untuk menghormati perbedaan agama.
8.	Saya peka terhadap perbedaan perilaku berdasarkan budaya.
9.	Saya prihatin secara emosional terhadap ketidakadilan etnis.
10.	Saya peka terhadap perbedaan status ekonomi.
11.	Status sosial seseorang tidak mempengaruhi perlakuan saya terhadap orang lain.
Dimensi Tindakan (Act)	
12.	Saya tidak melakukan apa-apa untuk menghentikan keadilan berdasarkan etnis.
13.	Saya tidak melakukan apa-apa untuk secara aktif melawan prasangka beragama.
14.	Saya tidak melakukan apapun ketika menyaksikan perlakuan tidak adil berdasarkan orientasi seksual.

Pengumpulan data dilakukan dengan *purposive sampling*, sehingga didapatkan responden yang dapat memberikan informasi yang sesuai (Campbell et al., 2020). Peneliti melakukan survey dengan bantuan *google form* dengan waktu pengumpulan data 1 bulan,

yang juga bertujuan untuk mempermudah responden penelitian. Survey merupakan penelitian yang mengumpulkan data dari perspektif individual, kemudian digeneralisasikan ke kelompok besar, yaitu dari sampel ke populasi (Jones et al., 2013). Penelitian ini juga telah memenuhi persyaratan dari komite etik dan peneliti juga telah mengupayakan untuk meminta persetujuan peneliti asli sebelum melaksanakan penelitian ini.

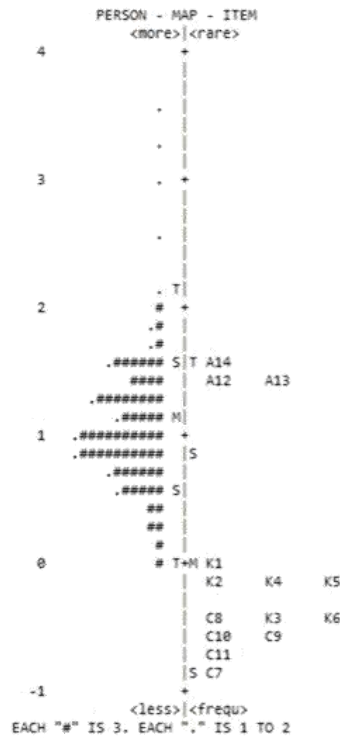
Analisis data menggunakan Model Rasch, yaitu metode pengukuran psikologis yang dapat membandingkan data respon dari alat ukur. Model Rasch menggunakan logit dalam menjawab item dan mampu memperkirakan kesalahan yang berkaitan dengan item. Analisis data dilakukan dengan aplikasi Winstep 5.1.0, dengan beberapa jenis analisis antara lain *unidimensionality*, *rating scale*, *item fit*, *wright map*, reliabilitas masing-masing item dan *person* serta interaksi *person-item*, validitas konstruk dengan melihat *percentage variance explained*, *infit* dan *outfit*, dan *Differential Item Functioning* (DIF).

Unidimensionality adalah memastikan bahwa item yang diukur hanya mengukur satu dimensi saja, seperti yang ditemukan dalam *factor loading*. Sedangkan *Wright map* merupakan grafik lokasi responden yang dibandingkan dengan lokasi item (Briggs, 2019). Reliabilitas pada Model Rasch didapatkan melalui 2 hal yaitu reliabilitas item dan reliabilitas *person*. Indeks reliabilitas item melihat apakah item yang sama akan memunculkan respon sama pada orang berbeda, sedangkan indeks reliabilitas *person* melihat apakah pergantian responden mempengaruhi konsistensi hal yang diukur masing-masing konstruk (Yasin et al., 2018).

Infit dan *outfit* dari *mean square* merupakan sebuah fitur pada Winstep yang difungsikan untuk melihat keakuratan item. Keakuratan item berarti butir perlu menjadi *fit* yang menandakan bahwa pemahaman orang yang mengerjakan terhadap butir sudah benar. Pada dasarnya *mean square* yang diharapkan adalah 1. Proses analisis nantinya juga menggambarkan *information function* dari setiap butir item (ICC). Yang terakhir adalah DIF, yang menunjukkan item yang berpotensi bias ($P < 0.05$) (Hagquist & Andrich, 2017)).

III. Hasil dan Pembahasan

3.1 Wright Map



Keterangan:

= Person ability
 (Kecenderungan orang untuk menampilkan agreement dalam suatu item)

K1 – K6
 Aspek Knowledge (Pengetahuan) Butir 1-6

C7 – C11
 Aspek Care (Kepedulian) Butir 7-11

A12 – A14
 Aspek Act (Tindakan) Butir 12-14

Interpretasi:

Seluruh tanda # berada hanya pada layer 0 hingga ke atas (tidak ada pada batas bawah) sehingga persebaran tidak merata (tidak semua

Grafik 1. Wright Map

Wright map adalah gambar yang menjelaskan posisi *item difficulty* (tingkat kesulitan soal) dan *person ability* (kemampuan seseorang dalam menjawab butir) (Abdullah et al., 2017). Dalam konteks penelitian ini ability mewakili agreement, dalam arti setiap item perlu memfasilitasi tingkatan kemampuan yang beragam. Di dalam *wright map* menggunakan logit yang menggambarkan posisi *respondent agreement* dan *item difficulty*. Penelitian terdahulu mengenai validitas dan reliabilitas MASQUE (Munroe, 2006; Permatasari et al, 2020) tidak menggambarkan posisi item dan partisipan dikarenakan *wright map* merupakan ciri khas dari Model Rasch.

Dalam bagian *wright map*, akan didapatkan informasi mengenai perbandingan kemampuan seseorang dan kesulitan item. Idealnya,. Namun, berdasarkan grafik *wright map*, hanya item 12, 13 dan 14 yang mampu mengukur abilitas responden yang tinggi dalam penelitian ini. Namun, sebagian besar responden dengan abilitas menengah tidak terfasilitasi karena tidak ada item yang berada pada tingkat kesulitan tersebut. Sebagian besar item justru berada di bagian bawah dan belum mampu memfasilitasi abilitas responden. Melalui hal ini, dapat diinterpretasikan bahwa sebagian besar butir memiliki tingkat kesulitan terlalu rendah, serta tidak ada butir untuk mengukur abilitas paling tinggi dari responden.

3.2 Unidimensionality

Tabel II. Raw Variance Explained by Measures

Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)			
		-- Empirical --	Modeled
Total raw variance in observations	=	27.1 100.0%	100.0%
Raw variance explained by measures	=	13.1 48.3%	50.7%
Raw variance explained by persons	=	2.5 9.4%	9.8%
Raw Variance explained by items	=	10.5 38.9%	40.8%
Raw unexplained variance (total)	=	14.0 51.7%	100.0% 49.3%
Unexplned variance in 1st contrast	=	4.1 15.2%	29.5%
Unexplned variance in 2nd contrast	=	1.6 5.9%	11.5%
Unexplned variance in 3rd contrast	=	1.3 4.7%	9.0%
Unexplned variance in 4th contrast	=	1.2 4.4%	8.5%
Unexplned variance in 5th contrast	=	1.0 3.8%	7.4%

Dalam Model Rasch, sebuah skala dapat disebut unidimensional (mengukur satu hal saja) bila *raw variance explained by measures* berada di atas 40% (Holster & Lake, 2016). *Raw variance explained by measures* yang didapatkan dalam analisis data sebesar 48,3%, yang menunjukkan bahwa skala MASQUE merupakan skala yang unidimensional. Unidimensi artinya bahwa dimensi karakter peserta yang diukur oleh suatu tes itu tunggal. Hasil ini bertentangan dengan penelitian terdahulu oleh Permatasari et al. (2020) yang menemukan adalah MASQUE adalah multidimensional. Hasil Model Rasch menunjukkan bahwa model ini dipersepsi oleh responden atau partisipan sebagai unidimensional.

3.3 Rating Scale Diagnostic

Tabel III. Rating Scale Diagnostic

Rating Scale	Skor	F	%	Average Measure	Infit MNSQ	Outfit MNSQ	Threshold
1	1	56	2	-.11	1.92	1.92	NONE
2	2	240	8	-.27	1.06	1.00	-2.00
3	3	240	8	.01	.82	.80	-.13
4	4	645	22	.81	.94	.93	-.41
5	5	990	34	1.34	1.29	1.21	.72
6	6	763	26	1.84	.90	.94	1.82

Berdasarkan *rating scale diagnostic*, *average measure* terus menunjukkan peningkatan yang berurutan pada setiap rating scale-nya. *Rating scale diagnostic* dilakukan untuk melihat apakah responden mampu mengidentifikasi perbedaan antara tiap opsi jawaban. Dengan terus menunjukkan peningkatan, berarti responden mampu membedakan di antara jawaban yang tersedia. *Rating scale diagnostic* juga merupakan hal yang khas dari Model Rasch yang menunjukkan apakah responden mampu membedakan respon terhadap skala yang digunakan. Penelitian terdahulu (Munroe, 2006; Permatasari et al., 2020) menggunakan analisis faktor, tidak menganalisis rating scale diagnostic.

3.4 Reliability

Tabel IV. Item Reliability

SUMMARY OF 14 MEASURED ITEM								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	955,9	210,0	.00	.08	.98	-.2	1.01	-.0
S.D.	171,2	.0	.00	.01	.32	3,2	.34	3,4
MAX.	1106,0	210,0	1,51	.09	1,58	5,9	1,63	6,1
MIN.	624,0	210,0	-.85	.06	.66	-3,6	.67	-3,6
REAL RMSE	.08	TRUE SD	.80	SEPARATION	9,71	ITEM	RELIABILITY	.99
MODEL RMSE	.08	TRUE SD	.80	SEPARATION	10,15	ITEM	RELIABILITY	.99
S.E. OF ITEM MEAN = .22								

UMEAN=.0000 USCALE=1.0000
 ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00
 2940 DATA POINTS, LOG-LIKELIHOOD CHI-SQUARE: 7234.93 with 2713 d.f. p=.0000
 Global Root-Mean-Square Residual (excluding extreme scores): .9242

Tabel V. Person Reliability

SUMMARY OF 210 MEASURED PERSON								
	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	63,7	14,0	1,08	.30	1,06	-.1	1,01	-.2
S.D.	5,6	.0	.53	.04	.63	1,6	.61	1,5
MAX.	81,0	14,0	3,60	.62	3,35	4,0	3,29	3,9
MIN.	49,0	14,0	-.06	.26	.03	-5,4	.04	-5,1
REAL RMSE	.34	TRUE SD	.41	SEPARATION	1,19	PERSON	RELIABILITY	.59
MODEL RMSE	.30	TRUE SD	.44	SEPARATION	1,45	PERSON	RELIABILITY	.68
S.E. OF PERSON MEAN = .04								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .61

Ada dua interpretasi yang bisa didapat dari bagian reliability, yaitu person reliability dan *item reliability*. Yang pertama adalah *person reliability*, yaitu orang yang terlibat dalam pengisian skala dapat mengisi dengan konsisten apabila angka person reliability berada di atas 0,8. Dalam hal ini, hasil analisis Rasch menunjukkan bahwa dalam mengisi, responden masih kurang konsisten karena angka *person reliability* hanya berada pada 0,59. Selain itu, *person separation* juga menunjukkan angka yang cukup rendah, yaitu 1,19 yang berarti butir kurang sensitif untuk menangkap keberagaman kemampuan responden. Namun sebaliknya, *item reliability* menunjukkan angka reliabilitas yang sangat baik, yaitu 0,99. *Item separation* juga menunjukkan angka yang baik yaitu 9,71, yang berarti responden cukup bervariasi untuk mengidentifikasi keberagaman item.

Di dalam Model Rasch terdapat person reliabilitas, item reliabilitas dan interaksi person-item-reliabilitas-alpha cronbach. Penelitian terdahulu menunjukkan reliabilitas yang cukup baik (0.753), sedangkan dalam penelitian ini Alpha Cronbach hanya 0.61. Namun, pada item reliability menunjukkan reliabilitas yang sangat tinggi. Hasil ini menunjukkan bahwa reliabilitas item sudah bagus, namun kurang pada person. Hal ini sejalan dengan hasil

pada wright map yang menunjukkan kurangnya item untuk mengukur orang dengan tingkat agreement yang rendah yaitu pada item 5, 6, 12, 13, dan 14.

3.5 Item Fit

Tabel VI. Item Fit

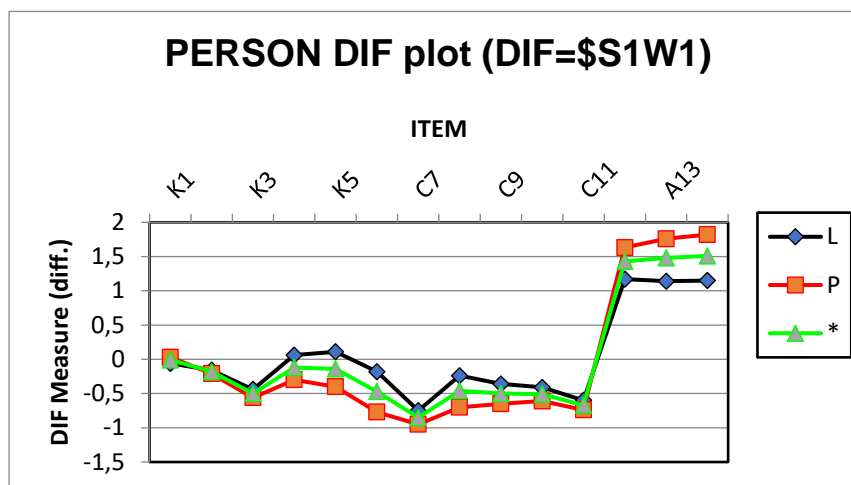
Dimensi - Item	Infit		Outfit		PT-MEASURE CORR.
	MNSQ	ZSTD	MNSQ	ZSTD	
Act – 14	1.58	5.7	1.63	6.0	.07
Act – 12	1.58	5.9	1.62	6.1	.13
Act – 13	1.49	5.0	1.53	5.2	.12
Care – 11	1.14	1.3	1.32	2.7	.35
Know – 4	.95	-.5	1.00	.1	.45
Know – 5	.89	-1.1	.93	-.7	.51
Know – 6	.86	-1.3	.86	-1.3	.55
Care – 10	.82	-1.8	.81	-1.9	.54
Care – 9	.80	-2.0	.80	-2.0	.55
Care – 8	.80	-2.0	.79	-2.1	.56
Care – 7	.78	-2.1	.77	-2.2	.52
Know – 1	.72	-3.0	.71	-3.1	.57
Know – 2	.68	-3.4	.67	-3.6	.59
Know – 3	.66	-3.6	.68	-3.3	.47

Selanjutnya, Model Rasch mampu mengidentifikasi *item fit*, yaitu sejauh mana item cocok menjadi bagian dari skala yang mengukur *multicultural attitude*. *Item fit* merupakan analisis khas dari Model Rasch mengenai batas-batas psikometri property yang ditolerir oleh model rasch. Pada penelitian terdahulu (Munroe, 2006; Permatasari et al., 2020) tidak menggunakan Model Rasch sehingga tidak bisa dibandingkan. Beberapa komponen penting dalam item fit adalah *infit*, *outfit*, *mean-square fit*, dan *standardized fit statistics (Zstd)*. Infit adalah pola respon item terhadap target subjek, sedangkan outfit adalah respon item terhadap subjek yang sebenarnya bukan targetnya. Sedangkan *mean-square* dapat melihat prediktabilitas item, karena sebuah item idealnya melakukan pengukuran acak dan tidak terlalu berulang atau mudah ditebak. Nilai ideal dari mean-square adalah sedekat mungkin dengan 1.0. Yang terakhir adalah *standardized fit statistics (Zstd)* yang mengukur apakah data fit terhadap model dengan nilai ideal 0,0. Bila lebih menunjukkan prediktabilitas tinggi, sedangkan bila kurang menunjukkan sebaliknya (Azizah & Wahyuningsih, 2020).

Mean-square value untuk *infit* maupun *outfit* yang ideal adalah antar 0.5 hingga 1.5, sedangkan nilai *point measure correlation* yang baik adalah antara 0.4 hingga 0.85. Berdasarkan *item fit*, ditemukan bahwa item 14 dan 12 memiliki infit di atas 1.5, sedangkan item 14, 12 dan 13 memiliki outfit di atas 1,5. Hal ini menunjukkan bahwa ketiga item tersebut tidak fit untuk mengukur *multicultural attitude*. Hal lain yang ditemukan dari bagian ini adalah item 11,12,13 dan 14 memerlukan perbaikan karena memiliki *point measure*

correlation dibawah 0,4. Berdasarkan *Zstd*, seluruh item cenderung menunjukkan prediktabilitas, kecuali item nomor 14 yang mendekati 0.0 ($n = 0.07$).

3.6 Differential Item Functioning (DIF)



Grafik 1. Differential Item Functioning (DIF)

DIF merupakan analisis tambahan dari Model Rasch yang menunjukkan orang pada *agreement* yang setingkat (kecenderungan cara menjawab yang sama) (Hagquist & Andrich, 2017). Pada hasil di atas menunjukkan adanya beberapa item yang mengalami DIF. Adanya DIF ini dapat mengakibatkan konsekuensi perbedaan norma antara laki-laki dan perempuan pada item-item tertentu.

Ada beberapa indikator yang bisa dilihat, antara lain *Welch Probability* ($p < 0.05$), DIF Contrast ($p > 0.64$) dan Mantel-Haenszel Chi-square (> 1.50). Berdasarkan hasil analisis, didapatkan bahwa tidak terlalu banyak item yang terindikasi DIF, namun berdasarkan tiga indikator didapatkan hasil yang berbeda. Hal ini berarti, butir yang terindikasi DIF berdasarkan *Welch Probability* belum tentu terindikasi oleh DIF Contrast dan Mantel-Haenszel Chi-square nya. Berdasarkan *Welch Probability*, item 5, 6, 12, 13, dan 14 terindikasi DIF. Berdasarkan DIF Contrast, hanya item 14 yang terindikasi DIF. Sedangkan, berdasarkan Mantel-Haenszel Chi-square, item 1, 4, 5, 6, 8, 12, 13, dan 14 terindikasi DIF. Hal ini menunjukkan item-item yang mungkin memiliki bias gender sehingga perlu penyesuaian atau perbaikan item.

IV. Simpulan dan Saran

Berdasarkan seluruh rangkaian analisis dengan Model Rasch, diperoleh berbagai data dengan berbagai perspektif. Menurut analisis Model Rasch, MASQUE versi Indonesia hanya mengukur satu dimensi padahal sesungguhnya MASQUE versi Indonesia diciptakan dengan

tiga dimensi sehingga dapat dikaji lebih jauh. Selain itu, responden MASQUE versi Indonesia tidak selaras dengan keberagaman kesulitan item yang ditemukan melalui Wright Map dan Person Separation, namun memiliki reliabilitas item dan item separation yang baik. Selain itu, beberapa item dalam MASQUE versi Indonesia semakin perlu untuk dikaji ulang karena ditemukan DIF dalam beberapa itemnya. Kesimpulannya, instrument MASQUE versi Bahasa Indonesia memiliki butir-butir yang sudah fit dan cukup bervariasi tingkat kesulitan itemnya, namun perlu dilakukan penyesuaian agar dapat memfasilitasi abilitas person yang lebih bervariasi dan item yang tersebar dalam aspek lebih sesuai.

Daftar Pustaka

- Abdullah, N., Noranee, S., & Khamis, M. R. (2017). The use of Rasch wright map in assessing conceptual understanding of electricity. *Pertanika Journal of Social Sciences and Humanities*, 25(S), 81–88.
- Azizah, A., & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *JUPITEK: Jurnal Pendidikan Matematika*, 3(1), 45–50. <https://doi.org/10.30598/jupitekvol3iss1pp45-50>
- Briggs, D. C. (2019). Interpreting and visualizing the unit of measurement in the Rasch Model. *Measurement: Journal of the International Measurement Confederation*, 146, 961–971. <https://doi.org/10.1016/j.measurement.2019.07.035>
- Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: complex or simple? Research case examples. *Journal of Research in Nursing*, 25(8), 652–661. <https://doi.org/10.1177/1744987120927206>
- Cobb, C. L., Lilienfield, S. O., Schwartz, S. J., Frisby, C., & Sanders, G. L. (2020). Rethinking Multiculturalism: Toward a Balanced Approach. *American Journal of Psychology*, 133(3), 275–293. <https://doi.org/10.5406/amerjpsyc.133.3.0275>
- Gebeyehu, D. (2022). A Focus on Banks' Multicultural Education and Global Citizens. *Academia Letters*, January, 1–3. <https://doi.org/10.20935/al4529>
- Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes*, 15(1), 1–9. <https://doi.org/10.1186/s12955-017-0755-0>

- Holster, T. A., & Lake, J. (2016). Guessing and the Rasch Model. *Language Assessment Quarterly*, 13(2), 124–141. <https://doi.org/10.1080/15434303.2016.1160096>
- Hoon, C. Y. (2017). Putting Religion into Multiculturalism: Conceptualising Religious Multiculturalism in Indonesia. *Asian Studies Review*, 41(3), 476–493. <https://doi.org/10.1080/10357823.2017.1334761>
- Munroe, A., & Pearson, C. (2006). The Munroe Multicultural Attitude Scale Questionnaire: A new instrument for multicultural studies. *Educational and Psychological Measurement*, 66(5), 819–834. <https://doi.org/10.1177/0013164405285542>
- Parker, L. (2017). Intersections of Gender/Sex, Multiculturalism and Religion: Young Muslim Minority Women in Contemporary Bali. *Asian Studies Review*, 41(3), 441–458. <https://doi.org/10.1080/10357823.2017.1332004>
- Parker, L. (2014). Religious education for peaceful coexistence in Indonesia? *South East Asia Research*, 22(4), 487–504. <https://doi.org/10.5367/sear.2014.0231>
- Permatasari, I., Milla, M. N., Lestari, S., Yusya, N., Adira, N., & Baswara, B. (2020). Adaptasi alat ukur Munroe Multicultural Attitude Scale Questionnaire versi Indonesia. *Jurnal Psikologi Sosial*, 18(2), 169–178. <https://doi.org/10.7454/jps.2020.17>
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61–70. <https://doi.org/10.32923/tarbawy.v7i1.1187>
- Setiawati, F. A., Amelia, R. N., Sumintono, B., & Purwanta, E. (2023). Study Item Parameters of Classical and Modern Theory of Differential Aptitude Test: Is it Comparable? *European Journal of Educational Research*, 12(2), 1097–1107. <https://doi.org/10.12973/eu-jer.12.2.1097>
- Yasin, S. N. T. M., Yunus, M. F. M., & Ismail, I. (2018). The use of rasch measurement model for the validity and reliability. *Journal of Counseling and Educational Technology*, 1(2), 22. <https://doi.org/10.32698/0111>
- Zarbaliyev, H. (2017). Multiculturalism in Globalization Era: History and Challenge for Indonesia. *Journal of Social Studies (JSS)*, 13(1), 1–16. <https://doi.org/10.21831/jss.v13i1.16966>