

---

## Pengaplikasian Multiple Indicator Multiple Causes (MIMIC) Model dalam Mendeteksi Differential Item Functioning (DIF) pada Alat Ukur Social Quality of Life

Hasbi Wahyudi

Universitas Islam Riau, Indonesia

[hasbiwahyudi76@gmail.com](mailto:hasbiwahyudi76@gmail.com)

### Abstract

*This study aims to detect DIF (differential item functioning) on a quality of life measurement tool that measures one aspect, namely social quality of life. Social quality of life contains 24 items developed from the Patient Reported Outcomes Measurement Information System (PROMIS) by a National Institutes of Health (NIH). This measuring tool measures the quality of life in the social function domain of adolescent patients suffering from diseases or chronic medical conditions. Detection of DIF in this study uses a special case approach from CFA, namely CFA with covariate or multiple indicator multiple causes (MIMIC) models. This study involved 322 participants, 117 (36%) male participants and 205 (64%) female participants, with an age range between 13-23 years in Riau Province. Based on the results of the first order CFA on a set of social quality of life items there are 22 valid items. Then the MIMIC model analysis results found that the model is fit with data where the value of RMSEA = 0.048, so it is known two items that contain DIF, namely item 5 (0.135,  $P = 0.002$ ) "I have a close friend" and item 23 (0.308,  $P = 0.002$ ) "I hope to have lots of friends".*

**Keywords:** *Social quality of life; MIMIC model; differential item functioning*

### Abstrak

Penelitian ini bertujuan untuk mendeteksi DIF (*differential item functioning*) pada alat ukur quality of life yang mengukur salah satu aspek yaitu *social quality of life*. *Social quality of life* berisi 24 item yang dikembangkan dari *Patient Reported Outcomes Measurement Information System* (PROMIS) oleh sebuah badan *National Institutes of Health* (NIH). Alalat ukur ini mengukur kualitas hidup pada domain fungsi sosial pasien remaja yang menderita penyakit atau kondisi medis kronis. Pendeteksian DIF pada penelitian ini menggunakan pendekatan kasus khusus dari CFA, yakni CFA with covariate atau multiple indicator multiple causes (MIMIC) model. Penelitian ini melibatkan 322 partisipan, yakni sebanyak 117 (36%) partisipan laki-laki dan 205 (64%) partisipan perempuan, dengan rentang usia antara 13-23 tahun di Propinsi Riau. Berdasarkan hasil first order CFA pada sekumpulan item-item social quality of life terdapat 22 item yang valid. Kemudian hasil analisis model MIMIC ditemukan bahwa model fit dengan data dimana nilai RMSEA = 0.048, sehingga diketahui dua item yang mengandung DIF, yaitu item 5 (0.135,  $P = 0.002$ ) "saya memiliki teman dekat" dan item 23 (0.308,  $P = 0.002$ ) "saya berharap mempunyai banyak teman".

**Kata kunci:** *Social quality of life; MIMIC model; differential item functioning*

## Pendahuluan

Konstruk atau atribut yang diukur dalam bidang psikologi dan pendidikan banyak yang bersifat laten (*latent variable*). Variabel laten merupakan konstruk yang tidak dapat diamati atau diukur secara langsung. Untuk mengukur sebuah variabel laten diperlukan pendefinisian operasional dari konstruk atau atribut yang diukur. Proses pendefinisian operasional dilakukan dengan menentukan dimensi, indikator dan penyusunan item dari konstruk atau atribut yang diukur (Fox, 2010). Pengukuran bertujuan untuk memperoleh skor yang mencerminkan karakteristik dari individu (Allen & Yen, 1979). Skor dapat diperoleh dari respon peserta tes terhadap pertanyaan atau pernyataan yang mewakili sebuah konstruk. Dimana skor yang diperoleh dalam pengukuran terhadap sampel perilaku akan dapat dievaluasi dan dinilai dengan prosedur standar sebagai acuan dalam mengambil sebuah keputusan (Anastasi & Urbina, 2004).

Tes merupakan salah satu alat pengukuran yang paling sering digunakan pada bidang psikologi dan pendidikan. Menurut Kaplan dan Saccuzo (2005) terdapat dua jenis tes psikologi, yaitu tes kepribadian (*personality test*) dan tes kemampuan (*ability test*). Penggunaan tes sebagai alat ukur sering dihadapi pada sebuah hambatan, yaitu kerentanan terhadap heterogenitas populasi (Sills & Brown, 2006). Heterogenitas populasi yang dimaksud adalah adanya perbedaan karakteristik individu, seperti latar belakang demografis, status sosial, status kesehatan maupun jenis kelamin. Tes yang rentan terhadap heterogenitas populasi mengakibatkan peserta tes dapat memberikan respon atau informasi yang bias. Pada pelaksanaannya, tes seharusnya berasaskan objektif, transparan, akuntabel dan tidak diskriminatif.

Sejak tahun 1960 terdapat persyaratan tambahan agar instrumen tes dikatakan baik, yaitu butir soal harus bersifat adil (*fairness*), dimana tes tidak mengandung item atau butir yang bias (Cheng, Shao & Lathrop, 2016). Item atau butir dikatakan bias ketika mengandung isi atau bahasa yang menguntungkan atau merugikan subkelompok tertentu. Ketika instrumen tes yang digunakan menguntungkan atau merugikan subkelompok tertentu, maka tes tersebut memuat bias butir atau mengandung *differential item functioning* atau yang disingkat dengan DIF. DIF adalah suatu karakteristik butir yang dapat menyebabkan probabilitas responden yang berkemampuan sama namun menghasilkan probabilitas menjawab benar yang berbeda. Misalnya, sulit atau mudahnya item dipengaruhi oleh berbagai faktor, seperti etnik, budaya, bahasa, agama atau jenis kelamin subkelompok (Tutz & Schaunberger, 2013). DIF menjadi penting untuk diuji jika terdapat keraguan dua kelompok tidak akan mendapatkan perlakuan yang adil meski mendapat stimulus yaitu berupa aitem yang sama.

DIF dijelaskan sebagai keadaan di mana dua individu kemampuan sama namun memiliki probabilitas yang berbeda untuk menjawab pertanyaan atau pernyataan dari sebuah tes. Tes yang memuat bias butir atau mengandung DIF akan mengakibatkan skor tes yang diperoleh tidak akan bermakna, dimana skor tes yang dihasilkan tidak menggambarkan karakteristik yang sebenarnya dari responden. Oleh karena itu, sangat penting untuk diperhatikan bagaimana akibat yang ditimbulkan oleh hasil tes yang mengandung bias butir, karena tidak hanya akan merugikan peserta tes, melainkan juga orang-orang lain yang ikut berkepentingan (*social consequences of test*).

Metode-metode psikometri dapat dipergunakan sebagai alat untuk mencapai dua jenis tujuan, diantaranya untuk mendapatkan persamaan matematis yang paling handal dalam meramalkan akibat dari suatu keputusan yang akan diambil berdasarkan skor dari satu atau sehimpunan tes, dan untuk menguji apakah suatu model teoritis tentang cara penggunaan skor tes untuk tujuan tertentu yang selama ini mungkin telah sering dipakai yang memang cukup handal dan dipercaya (Umar, 2011).

Beberapa peneliti terdahulu telah melakukan pendeteksian DIF pada sebuah tes dengan berbagai macam cara atau metode, seperti penggunaan *item response theory* (Masitah, 2012), regresi logistik (*logistic regression*) (Scot *et al.*, 2010) dan penggunaan *multiple indicator multiple causes* (MIMIC) model (Mucherah, Finch & Keaikitse, 2012). Dengan pesatnya perkembangan teori pengukuran maka teknik-teknik

baru dalam mengevaluasi properti psikometris hasil pengukuran mulai bermunculan. Salah satunya adalah melibatkan kovariat yang berfungsi sebagai variabel penyerta dalam *confirmatory factor analysis* (CFA). Model CFA dengan melibatkan kovariat atau yang dikenal dengan *multiple indicator multiple causes* (MIMIC) *model* atau yang disingkat dengan model MIMIC merupakan sebuah kasus khusus dari model CFA umum yang disertai dengan kovariat, dimana berisi variabel teramati yang memiliki *Multiple Indicator* dan *Multiple Causes* dari sebuah variabel laten (Bollen, 1989).

Model MIMIC merupakan metode yang paling populer digunakan untuk mendeteksi DIF pada penelitian-penelitian terbaru, seperti penelitian yang dilakukan oleh Finch (2005), Shih & Yang (2009), Mucherah *et al.* (2011), Finch & French (2011), Watters (2013) dan Cheng, Shao & Lathrop, (2016). Hal ini dikarenakan penggunaan model MIMIC cenderung lebih praktis dalam mendeteksi DIF (Gallo, Anthony & Muthen, 1994) dan model MIMIC juga telah terbukti efektif untuk mengidentifikasi DIF (Finch, 2005; Woods, 2009). Dimana model MIMIC tidak perlu melakukan analisis pada masing-masing kelompok, seperti yang dilakukan pada *multiple group* CFA. Selain itu, model MIMIC juga memungkinkan variabel lain baik kategori maupun kontinum dapat diikutsertakan dalam analisis (Glockner & Hoijtink, 2003).

Penelitian ini akan menggunakan model MIMIC dalam mendeteksi DIF pada alat ukur *quality of life* (QoL). *Quality of life* di berbagai bidang atau disiplin ilmu seperti lingkungan, kesehatan, masyarakat, ekonomi maupun kedokteran memiliki definisi tersendiri sesuai pada pendekatan ataupun tujuan pengukurannya. Menurut Sarafino dan Smith (2011) *quality of life* merupakan hal yang penting dalam kepedulian seseorang tentang kesehatannya, karena *quality of life* akan berkurang apabila seseorang mengalami sakit. *Quality of life* menjadi pertimbangan yang penting untuk melakukan usaha pencegahan pada kondisi seseorang baik sebelum ataupun sesudah mengalami penyakit. Selain itu, *World Health Organization* (WHO) mendefinisikan *quality of life* sebagai kondisi yang berdasarkan persepsi individu dalam kehidupan pada konteks sistem nilai dan budaya dimana mereka tinggal dan berdasarkan kaitannya dengan tujuan hidup masing-masing individu, harapan, standar dan kepentingannya.

*Quality of life* merupakan konstruk yang lebih luas dari *subjective well-being*. Karena *quality of life* merupakan hal kompleks yang mencakup berbagai jenis domain yang terkait dengan status kesehatan seperti aktivitas yang dilakukan sehari-hari, status pekerjaan dan fungsi-fungsi sosial dalam pertemanan ataupun hubungan lainnya. Beberapa peneliti menjelaskan bahwasanya *quality of life* berkaitan dengan karakteristik dari sosio-demografis, misalnya penelitian yang dilakukan oleh Trompenaars *et al.* (2005) menunjukkan bahwasanya usia memiliki hubungan yang negatif dengan kesehatan fisik dan *quality of life* untuk domain hubungan sosial. Selain itu, pada umumnya peneliti menyimpulkan bahwasanya perempuan memiliki *quality of life* yang lebih rendah jika dibandingkan dengan laki-laki (Sabbah *et al.*, 2003; Trompenaars *et al.*, 2005; Mankar *et al.*, 2010; Cruz *et al.*, 2011). Penelitian mengenai *quality of life* telah banyak melahirkan jenis instrumen atau tes dalam mengukur *quality of life*, diantaranya kuesioner *health related quality of life* (HRQL), *short form* (SF) 12, *short form* (SF) 36 dan WHOQOL BREF serta *social quality of life scale*, dimana alat ukur ini mengukur kualitas hidup pasien yang mengalami sakit menahun pada domain fungsi sosialnya (O'Connor 1993; Reeve *et al.*, 2007; Yousefy *et al.*, 2010; Cai 2010).

Penelitian ini berfokus pada penggunaan alat ukur *social quality of life* (QoL) yang dikembangkan untuk mengukur kualitas hidup pasien yang mengalami sakit menahun pada domain fungsi sosialnya. Alat ukur ini memiliki 24 item dari dua dimensi, yakni *positive experience* dan *negative experience* yang merupakan pengembangan dari *Patient Reported Outcomes Measurement Information System* (PROMIS; Reeve *et al.*, 2007), oleh sebuah badan *National Institutes of Health* (NIH). Pengembangan alat ukur ini dilakukan dalam membantu mencapai tujuan para dokter spesialis anak untuk mengetahui berbagai aspek kualitas interaksi sosial pasien remaja dengan rekannya, orang tua, guru dan sebagainya.

## Quality of Life

Liu mengatakan bahwa tiap-tiap individu memiliki definisi yang berbeda mengenai *quality of life* (dalam Felce & Perry, 2005). *Quality of life* merupakan sebuah konsep yang bersifat sangat subjektif, kemudian sifat subjektif dari *quality of life* ini membuat konseptualisasi dari kualitas hidup bervariasi antara satu peneliti dengan peneliti yang lain. Liu (dalam Felce & Perry, 2005) lebih lanjut menjelaskan bahwa secara umum *quality of life* menggambarkan kesejahteraan individual terhadap kehidupannya, namun sulit mendapatkan konsensus dalam mendefinisikan *quality of life* secara operasional. Hal senada juga diungkapkan oleh Molnar (2009) mengatakan bahwa pada dasarnya menyusun konsep mengenai *quality of life* adalah hal yang sulit.

Para ahli berpendapat bahwa konsep dan pengukuran *quality of life* harus berpusat pada persepsi subjektif individu mengenai kehidupannya sendiri (Mendlowicz & Mauro, 2000). Dimana *quality of life* yang berpusat pada subjektif memiliki kekuatan prediktif yang lebih tinggi dari pada *quality of life* yang berpusat pada objektif (Ruggeri, Warner, Bisoffi & Fontecendo, 2011). Carr dan Higginson (2006) mengungkapkan bahwa *quality of life* merupakan suatu konstruk yang bersifat individual. Sehingga, komponen objektif dari QoL tidak mempengaruhi QoL itu sendiri secara langsung melainkan diperantarai oleh persepsi individu. *Quality of life* merupakan interaksi antara penghayatan subjektif dan bobot kepentingan dalam aspek-aspek kehidupan tertentu, dengan beberapa faktor kondisi kehidupan yang dapat berpengaruh pada persepsi individu mengenai berbagai kondisi kehidupan.

Moons, Marquet, Budst & de Geest (2004) mempermudah konseptualisasi mengenai *quality of life*, dengan memaparkan hal-hal penting dalam konseptualisasi *quality of life*, yaitu kualitas hidup tidak boleh disamakan dengan status kesehatan ataupun kemampuan fungsional, kualitas hidup didasarkan oleh evaluasi subjektif daripada parameter objektif, tidak terdapat perbedaan yang jelas antara indikator-indikator kualitas hidup dengan faktor yang menentukan kualitas hidup. *Quality of life* dapat berubah seiring waktu dan dapat dipengaruhi secara positif maupun negatif. Moons *et al* (2004) mendefinisikan kualitas hidup sebagai berikut:

*“The degree of overall life satisfaction that is positively or negative influenced by individual perception of certain aspect of life important to them...”*

World Health Organization (WHO) mendefinisikan *quality of life* sebagai persepsi individu mengenai posisi mereka dalam kehidupan dilihat dari konteks budaya dan sistem nilai dimana mereka tinggal serta hubungannya dengan tujuan, harapan, standar dan hal-hal yang menjadi perhatian individu (Power, 2013). Definisi *quality of life* dalam definisi WHO ini menekankan adanya persepsi individu mengenai posisi kehidupan yang dipengaruhi oleh budaya dan sistem nilai dimana individu tinggal. Berdasarkan beberapa definisi yang telah dijelaskan, peneliti memberi kesimpulan bahwa kualitas hidup adalah penilaian subjektif individu mengenai posisi kehidupannya pada beberapa aspek kehidupan yang penting baginya, dimana penilaian subjektif dapat dipengaruhi oleh sosiodemografis individu. Menurut World Health Organization (Power, 2013) kualitas hidup meliputi empat aspek, yaitu: (1) Domain fisik yang terdiri dari kenyamanan fisik dalam beraktivitas, tenaga yang dimiliki dan perasaan lelah, kesempatan untuk tidur dan istirahat. (2) Domain psikologis yang terdiri dari perasaan positif, kemampuan berfikir dan belajar saat menghadapi masalah, kemampuan mengingat dan konsentrasi dalam menjalankan usaha, harga diri, gambaran dan penampilan diri. (3) Domain hubungan sosial yang terdiri dari hubungan perorangan, dukungan sosial maupun aktivitas seksual. (4) Domain lingkungan yang terdiri dari keamanan lingkungan rumah, sumber penghasilan, kesehatan dan perhatian sosial, kesempatan untuk memperoleh informasi baru, partisipasi dalam kesempatan berekreasi dan waktu luang.

## Differential Item Functioning (DIF)

Sejak tahun 1960-an, terdapat persyaratan tambahan agar item dikatakan baik, yaitu butir soal harus adil (*fairness*). Pengujian untuk melihat apakah butir soal bertindak adil atau tidak disebut pengujian *differential item functioning* (DIF). Secara konseptual, DIF dikatakan muncul pada sebuah item jika peserta tes yang mempunyai kemampuan yang sama pada konstruk yang diukur oleh tes dan berasal dari kelompok berbeda, mempunyai probabilitas berbeda dalam menjawab item tersebut (Hulin, Drasgow & Parson, 1993; Woods, Oltmanns & Turkheimer 2009; dan Tutz & Schaunberger, 2013).

*Differential item functioning* (DIF) dapat dibedakan menjadi dua jenis, yaitu *uniform* DIF dan *non-uniform* DIF (Scott et al., 2010). *Uniform* DIF terjadi apabila sebuah item menunjukkan besaran DIF yang sama, meskipun besaran pada level  $\theta$  berbeda. Sedangkan *non-uniform* DIF menunjukkan besaran DIF yang berbeda dites pada  $\theta$  yang berbeda. Prosedur pendeteksian harus berusaha untuk menilai keduanya, baik *uniform* DIF maupun *non-uniform* DIF, walaupun dalam prakteknya tidak semua metode dapat mendeteksi *non-uniform* DIF (Scott et al., 2010).

Meskipun pada dasarnya DIF dan bias adalah sama, namun beberapa tokoh mencoba membedakan antara DIF dan bias. Dimana *differential item functioning (DIF)* merupakan bias yang terjadi pada item atau pernyataan. Sedangkan bias tes lebih kepada bias yang ada pada tes atau kumpulan-kumpulan dari item-item tersebut, yang sering disebut dengan *differential test functioning (DTF)* (Camilli dan Shepard, 1994).

Sedangkan berdasarkan titik kajian, maka bias dapat dibagi menjadi dua, yaitu eksternal dan internal. Jika bias tersebut dikaji pada seperangkat tes maka dapat dikatakan sebagai bias eksternal (Camilli dan Shepard, 1994). Jika terjadi ketidakadilan dalam suatu alat tes terhadap *testee* yang memiliki kemampuan yang sama pada kelompok yang berbeda, maka hal itulah disebut dengan bias eksternal. Sehingga fokus bias eksternal terletak pada validitas prediktif. Selain bias eksternal, terdapat juga bias internal. Menurut Camilli dan Shepard (1994) bias internal berkaitan dengan pengembangan tes. Bias item karena faktor internal dapat terjadi apabila kajian difokuskan pada komponen item (bentuk item, kalimat dan kata-yang digunakan, gambar maupun petunjuk).

## Metode

### Desain dan Sampel

Penelitian ini merupakan penelitian non-eksperimen dengan pendekatan secara kuantitatif. Hal ini dikarenakan penelitian ini tidak melakukan manipulasi dan mengkondisikan variabel serta tidak memberi perlakuan terhadap partisipan penelitian ini, sehingga data yang diperoleh dari partisipan berdasarkan *ex post facto* (Simon & Goes, 2013). Adapun pengambilan sampel dalam penelitian ini dilakukan dengan metode *non probability sampling* dengan teknik *convenience sampling*, dimana pengambilan sampel disesuaikan dengan maksud dan tujuan dari penelitian ini. Adapun yang menjadi sampel dalam penelitian ini adalah pasien remaja yang memiliki riwayat penyakit atau menderita medis kronis di Provinsi Riau yang berusia antara 13-23 tahun berjumlah 322 orang.

### Alat Ukur

Skala dalam alat ukur ini terdiri dari item-item yang memiliki empat pilihan respon mulai dari "Sangat Tidak Setuju" sampai "Sangat Setuju". Skala tersebut dipilih agar partisipan memberikan respon yang lebih bervariasi dalam rentang tertentu dan dapat mengevaluasi pernyataan yang ada sesuai dengan kondisi dirinya. Instrumen pengumpulan data dalam penelitian ini memiliki pernyataan positif dan negatif. Skor tertinggi diberikan pada pilihan jawaban sangat setuju dan skor terendah diberikan pada pilihan jawaban sangat tidak setuju untuk

pernyataan *favorable*, sedangkan pada pernyataan *unfavorable* skor tertinggi diberikan pada pilihan sangat tidak setuju dan skor terendah diberikan pada pilihan sangat setuju.

Berikut akan dijelaskan lebih rinci alat ukur *social quality of life* (QoL) oleh Cai (2010a, 2010b) yang telah melakukan standarisasi terhadap alat ukur *social quality of life*, dimana alat ukur ini mengukur kualitas hidup pasien anak-anak yang menderita kondisi medis kronis pada domain fungsi sosialnya. Alat ukur ini memiliki 24 item dari dua jenis, yakni *positive experience* dan *negative experience* yang merupakan pengembangan dari *Patient Reported Outcomes Measurement Information System* (PROMIS; Reeve et al., 2007), oleh sebuah badan *National Institutes of Health* (NIH), hal ini dilakukan sebagai bentuk mencapai tujuan para dokter anak untuk mengetahui berbagai aspek kualitas interaksi sosial pasien dengan rekannya, orang tua, guru dan sebagainya.

## MIMIC Model

Sebelum pendeteksian DIF pada alat ukur *social quality of life*, terlebih dahulu dilakukan pengujian validitas konstruk dalam penelitian ini menggunakan metode analisis faktor yaitu *confirmatory factor analysis* (CFA) model satu faktor (1<sup>st</sup> order) yang dilakukan untuk menguji unidimensionalitas alat ukur. Setelah mendapatkan item yang valid melalui uji validitas konstruk melalui CFA, kemudian peneliti akan melakukan pendeteksian DIF dengan menggunakan pendekatan model MIMIC pada sekumpulan item yang valid tersebut. Dimana terdapat kovariat yang akan diuji pengaruhnya terhadap item-item yang terbukti mengukur *social quality of life*. Adapun kovariat yang dimaksud dalam penelitian ini adalah jenis kelamin, yang terdiri dari dua kategori yaitu laki-laki yang diberi kode 0 dan perempuan yang diberi kode 1. Apabila *direct effect* dari jenis kelamin memberikan hasil yang signifikan, maka hal tersebut menunjukkan bahwa terjadi DIF pada item yang diuji karena terdapat perbedaan yang signifikan antara kelompok jenis kelamin laki-laki apabila dibandingkan dengan kelompok jenis kelamin perempuan. Sebagaimana telah dijelaskan pada bab sebelumnya, bahwa model MIMIC dapat didefinisikan sebagai berikut:

$$\begin{aligned}\eta &= \Gamma X + \zeta \\ Y &= U_y + \Lambda_y \eta + \varepsilon \\ X &\equiv \xi\end{aligned}$$

Dimana  $U_y$  adalah vektor *intercepts* dari indikator  $y_s$ . Pada persamaan (4), *intercepts* dari faktor tidak diikutsertakan, karena *factor means/intercept* dalam sebuah model *single group* harus dijadikan nol untuk tujuan dari identifikasi model. Perbedaan *factor mean* antara kelompok dapat diuji dalam model *multi-group*. Simbol  $\equiv$  menunjukkan identitas spesifik antara  $X$  dan  $\xi$ , yang diperoleh dengan memperbaiki *loading factor* ke 1,0 (yaitu,  $\Lambda_s = 1$ ) dan *measurement error* ke 0 (yaitu,  $\Theta_\delta = 0$ ) (Wang & Wang, 2012). Untuk menilai apakah model pengukuran benar-benar fit dengan data, maka perlu diperhatikan nilai indeks fit. Indeks fit yang dihasilkan dari analisis menggunakan metode CFA ada berbagai macam. Suatu indeks yang menunjukkan model tersebut *fit* tidak memberikan jaminan bahwa model benar-benar *fit*, begitupun sebaliknya. Oleh karena itu, dalam penelitian ini menggunakan kriteria indeks *fit* sebagai berikut:

### 1. Chi-Square

Nilai *chi-square* menunjukkan penyimpangan antara *sample covariance matrix* dengan *model covariance matrix* (Joreskog dan Sorbom, 1993). *Chi-square* merupakan ukuran mengenai buruknya *fit* suatu model. Probabilitas *chi-square* dalam pengujian ini diharapkan tidak signifikan ( $p > 0.05$ ) agar model dikatakan *fit* dengan data.

## 2. Root Mean Square Error of Approximation (RMSEA)

RMSEA diperkenalkan oleh Steiger dan Lind tahun 1980. RMSEA mengukur penyimpangan nilai parameter pada suatu model dengan matriks kovarians populasinya. Nilai RMSEA  $< 0.05$  mengindikasikan model *fit* dan nilai RMSEA yang berkisar antara 0.08 – 0.10 menyatakan bahwa model memiliki perkiraan kesalahan yang *reasonable*. Sementara nilai RMSEA 0.08 – 0.10 menunjukkan model memiliki *fit* yang cukup. Sedangkan nilai RMSEA  $> 0.1$  menunjukkan model *fit* yang sangat buruk. Tetapi, penggunaan RMSEA sangat sensitif dengan jumlah sampel yang digunakan. Semakin sedikit jumlah sampel maka akan menyebabkan *error* yang semakin membesar. Dengan demikian, maka dapat disimpulkan bahwasanya model *fit* dengan data ketika nilai RMSEA  $< 0.05$ .

## Hasil dan Pembahasan

Peneliti menguji apakah ke 24 item yang ada bersifat *unidimensional*, artinya benar hanya mengukur *social quality of life*. Dari hasil analisis CFA yang dilakukan dengan model satu faktor, ternyata tidak fit, dengan  $Chi-square = 500.255$ ,  $df = 252$ ,  $P-value = 0.0000$ , dan nilai RMSEA = 0.055, oleh sebab itu, peneliti melakukan modifikasi terhadap model, dimana kesalahan pengukuran pada beberapa item dibebaskan berkorelasi satu sama lainnya, maka diperoleh model fit dengan  $Chi-square = 413.186$ ,  $df = 247$ ,  $P-value = 0.0000$ , RMSEA = 0.046.

Setelah di dapat nilai RMSEA  $< 0.05$  dapat dinyatakan bahwa model dengan satu faktor dapat diterima. Artinya seluruh item hanya mengukur satu faktor yaitu *social quality of life*. Kemudian penulis melihat apakah item tersebut mengukur faktor yang hendak diukur secara signifikan dan sekaligus menentukan apakah item tersebut perlu didrop atau tidak, pengujiannya dilakukan dengan melihat nilai  $z$  bagi setiap koefisien muatan faktor.

Tabel I. Muatan Faktor CFA I<sup>st</sup> Order

Item	Estimate	Standard Error	z-value	Keterangan
1	0.392	0.059	6.655	Valid
2	0.547	0.046	11.773	Valid
3	0.454	0.061	7.463	Valid
4	0.004	0.072	0.062	Tidak Valid
5	0.341	0.060	5.660	Valid
6	0.392	0.066	5.961	Valid
7	0.290	0.070	4.172	Valid
8	0.313	0.071	4.405	Valid
9	0.249	0.066	3.743	Valid
10	0.277	0.086	3.213	Valid
11	0.502	0.056	8.971	Valid
12	0.570	0.051	11.195	Valid
13	0.421	0.066	6.419	Valid
14	0.516	0.071	7.244	Valid
15	0.228	0.074	3.073	Valid
16	0.452	0.067	6.799	Valid
17	0.210	0.071	2.975	Valid
18	0.584	0.054	10.899	Valid
19	0.681	0.046	14.883	Valid
20	0.417	0.060	6.9	Valid
21	0.697	0.040	17.608	Valid
22	0.194	0.066	2.934	Valid
23	0.238	0.056	4.241	Valid
24	-0.001	0.066	-0.010	Tidak Valid

Nilai  $z$  bagi koefisien muatan faktor pada 22 item signifikan karena  $z > 1.96$ , sedangkan dua item yaitu item 4 dan item 24 tidak valid karena tidak memenuhi kriteria. Selanjutnya penulis melihat muatan faktor dari

item, apakah ada yang bermuatan negatif atau tidak, lalu diketahui terdapat item yang muatan faktornya negatif, yaitu item 24. Sehingga pada analisis yang akan dilakukan untuk menguji model-model selanjutnya dalam penelitian ini item-item tersebut dikeluarkan, karena merupakan item yang tidak valid.

Setelah mendapatkan 22 item yang valid melalui pengujian *first order* CFA, didapat hasil bahwa ada dua item yang tidak valid yaitu item 4 dan 24. Selanjutnya peneliti akan melakukan pendeteksian DIF dengan menggunakan model MIMIC pada sekumpulan item yang valid tersebut. Pengujian dilakukan dengan cara melihat koefisien-koefisien yang akan dijelaskan selanjutnya.

Tabel 2. Hasil Estimasi Parameter Model MIMIC

Item	Estimate	Standard Error	z-value	Keterangan
1	0.389	0.059	6.590	Valid
2	0.545	0.046	11.769	Valid
3	0.452	0.061	7.433	Valid
5	0.288	0.063	4.555	Valid
6	0.394	0.065	6.053	Valid
7	0.299	0.070	4.288	Valid
8	0.315	0.071	4.444	Valid
9	0.252	0.066	3.818	Valid
10	0.275	0.087	3.182	Valid
11	0.499	0.056	8.891	Valid
12	0.563	0.051	10.985	Valid
13	0.425	0.065	6.508	Valid
14	0.515	0.071	7.292	Valid
15	0.235	0.075	3.154	Valid
16	0.453	0.066	6.852	Valid
17	0.210	0.071	2.973	Valid
18	0.588	0.054	10.895	Valid
19	0.681	0.045	15.189	Valid
20	0.421	0.060	7.079	Valid
21	0.697	0.040	17.581	Valid
22	0.216	0.066	3.299	Valid
23	0.129	0.061	2.107	Valid

Dari hasil analisis secara simultan, terdapat perbedaan mendasar dari model MIMIC apabila dibandingkan dengan dua model sebelumnya dimana pada model ini, terdapat kovariat yang akan diuji pengaruhnya terhadap item-item yang terbukti mengukur *social quality of life* dimana kovariat yang dimaksud adalah jenis kelamin yang terdiri dari dua kategori yaitu laki-laki yang diberi kode 0 dan perempuan yang diberi kode 1. Apabila *direct effect* dari jenis kelamin memberikan hasil yang signifikan, maka hal tersebut menunjukkan bahwa terjadi DIF pada item yang diuji karena terdapat perbedaan yang signifikan antara kelompok jenis kelamin laki-laki apabila dibandingkan dengan kelompok jenis kelamin perempuan.

Metode estimasi yang digunakan pada model MIMIC adalah metode estimasi *robust maximum likelihood* (MLR) yang akan menghasilkan hasil estimasi namun dengan indeks model fit yang *robust*. Adapun yang dimaksud dengan *robust* adalah metode estimasi ini tetap dapat berfungsi dengan baik meskipun terdapat asumsi dasar yang dilanggar, lalu metode ini menghasilkan indeks model *fit* yang jauh lebih baik dan terkoreksi.

Berdasarkan tabel di atas, didapat informasi signifikansi masing-masing item dari MIMIC model yang diteorikan mengukur satu faktor yaitu *social quality of life* sehingga setelah dilakukan pembuktian menggunakan model *first order* yang telah dijelaskan sebelumnya, *social quality of life* dapat dikonstantkan variansnya menjadi 1 untuk menguji apakah gender menjadi satu-satunya penyebab bervariasinya item-item yang ada dimana hal tersebut bertujuan untuk menguji DIF dimana hasil pengujiannya dijelaskan pada tabel di bawah ini:



Tabel 3. Pengujian DIF Dengan Model MIMIC Pada Alat Ukur Social Quality of Life

	Estimate	Std. Error	z-value	p-value
Quality of Life ON Gender $R^2 = 0.018, p=0.000$				
Quality of Life ON Gender	0.343	0.045	7.581	0.000
Item 5 ON Gender	0.137	0.052	2.623	0.009
Item 23 ON Gender	0.309	0.041	7.449	0.000

Berdasarkan tabel 3, terdapat 2 item yang mengandung DIF yaitu item 5 dan 23. Adapun kriteria pengujiannya dengan melihat hasil estimasi yang didapat dengan membuat model dimana terjadi *direct effect* dari *gender* pada masing-masing item. Hasil estimasi yang signifikan menunjukkan bahwa item tersebut berfungsi berbeda antara kelompok laki-laki dengan perempuan dengan kata lain item tersebut mengandung DIF. Lalu melihat pengaruh langsung dari jenis kelamin ke *latent variable Quality of Life* terdapat pengaruh yang signifikan dengan besaran  $R^2 = 0.018$  yang signifikan dan dengan koefisien regresi 0.343 yang signifikan dengan arah yang positif yang artinya terdapat perbedaan *mean* antara laki-laki dan perempuan dalam konteks merespon item *Quality of Life*.

Analisis mengenai item yang mengandung DIF tidak hanya berhenti sampai disini, satu persatu item yang mengandung DIF perlu dianalisis secara konten untuk mendapat gambaran tentang bagaimana item tersebut dapat berfungsi berbeda antara dua kelompok jenis kelamin yang berbeda. Berdasarkan tabel 3 *gender* memiliki pengaruh positif yang signifikan (0.137,  $p=0.009$ ) terhadap item 5 (*saya memiliki teman dekat*), ini mengimplikasikan bahwa perempuan cenderung memiliki skor lebih tinggi jika dibandingkan laki-laki. Selain itu, *gender* juga memiliki pengaruh positif yang signifikan (0.309  $p=0.000$ ) terhadap item 23 (*saya berharap mempunyai banyak teman*). Hal ini berarti bahwa partisipan perempuan cenderung memiliki skor lebih tinggi dibandingkan laki-laki.

## Pembahasan

Penelitian ini mengilustrasikan bagaimana model MIMIC dalam mendeteksi DIF pada salah satu alat ukur *quality of life*, yaitu *social quality of life*. *Social quality of life* merupakan alat ukur yang mengukur *quality of life* dalam domain fungsi sosial pada pasien remaja yang memiliki riwayat penyakit atau medis kronis. Hasil penelitian menunjukkan bahwasanya dari 22 item valid yang mengukur *social quality of life* terdapat dua item yang terdeteksi mengandung DIF yaitu item 5 "*saya memiliki teman dekat*" dan item 23 "*saya berharap mempunyai banyak teman*", kedua item tersebut memiliki signifikansi dengan arah yang positif mengimplikasikan bahwa partisipan perempuan cenderung memiliki skor lebih tinggi jika dibandingkan laki-laki dalam merespon item tersebut.

Jika kita simpulkan terhadap dua item yang terdeteksi DIF, dapat dikatakan bahwasanya kedua item tersebut berkaitan dengan relasi atau dukungan sosial yang diperoleh oleh individu ketika mengalami kesulitan atau hambatan dalam hidupnya. Sebagaimana telah dijelaskan oleh Taylor (2000) bahwasanya dalam merespon stress atau kesulitan yang dialami suatu individu, ternyata wanita mempunyai cara unik yang relatif tidak didapati pada pria, yakni dengan pola yang dinamakan "*tend and befriend*", suatu pola yang dibangun berdasarkan sistem "*caregiving*" melalui proses biobehavioral. Dengan menyalurkan naluri keibuannya, merawat anak-anak dan keluarga, atau melalui relasi sosial, terutama pada sesama wanita, mereka berusaha mencari dukungan untuk mereduksi stres atau kesulitan yang dialaminya.

Taylor (2000) lebih lanjut menjelaskan bahwasanya model "*tend and befriend*" muncul dan dipakai sebagai pendamping model "*fight or flight*" yang selama ini menjadi kajian respon individu ketika berhadapan dengan kesulitan atau stres. Berdasarkan hal ini, dimana adanya perbedaan antara laki-laki dan perempuan dalam menghadapi kesulitan atau stres yang dihadapi inilah yang menjadi asumsi penulis penyebab mengapa antara laki-laki dan perempuan merespon berbeda terhadap kedua item yang terdeteksi DIF.

Penelitian sebelumnya mengenai *quality of life* pada individu menyatakan bahwasanya ada perbedaan *quality of life* antara laki-laki dan perempuan (Sabbah, Drouby, Rude & Mercier, 2003; Trompenaars, Masthoff, Heck, Hodiattmont & Vries, 2005; Mankar, Joshi, Velankar, Mhatre & Nalgundwar, 2010; Cruz, Polanczyk, Camey, Hoffman & Fleck, 2011), hal ini menunjukkan bahwa item-item pada alat ukur yang mengukur *quality of life* rentan terhadap bias, sebagaimana telah dijelaskan bahwasanya alat ukur yang mengandung bias akan menghasilkan penelitian yang tidak akan bermakna. Oleh karena itu, berdasarkan hasil penelitian ini maka pada penelitian mendatang, *gender* merupakan variabel yang harus diperhitungkan dalam menganalisis data *social quality of life*. Penelitian ini juga mengukuhkan penggunaan alat ukur *social quality of life* yang sebelumnya telah dilakukan oleh Cai (2010a, 2010b) sekaligus melengkapinya dengan informasi lebih lanjut mengenai item-item yang mengandung DIF sehingga pada penelitian mendatang diharapkan para peneliti memasukkan variabel *gender* untuk melakukan koreksi terhadap faktor-faktor psikologis yang berperan terhadap bervariasinya *social quality of life*.

Hasil dalam penelitian ini juga menunjukkan bagaimana model MIMIC efektif dan praktis digunakan untuk mendeteksi DIF, karena estimasi parameter pada model pengukuran tidak dilakukan pada masing-masing kelompok. Akan tetapi kovariat dimasukkan kedalam model CFA, dimana kovariat dalam penelitian ini adalah variabel nominal yang menggambarkan tingkat kelompok yang diketahui (misalnya, jenis kelamin: 0 = laki-laki, 1 = perempuan). Hal ini sejalan dengan penelitian sebelumnya yang dilakukan oleh Gallo, Anthony dan Muthen (2004) yang menyatakan bahwasanya model MIMIC cenderung praktis digunakan dalam mendeteksi DIF yang spesifik dalam tingkat item. Selanjutnya, penelitian yang dilakukan oleh Finch (2005) dan Woods (2009) juga membuktikan hal yang sama, bagaimana MIMIC model menjadi salah satu cara atau metode yang efektif untuk mendeteksi DIF pada tingkat item.

Penelitian ini menggunakan ukuran sampel sebanyak 322 orang untuk mengoptimalkan tahapan estimasi dalam penelitian ini, berdasarkan hasil analisis pendeteksian DIF dengan model MIMIC mampu menghasilkan model yang baik. Hal ini sejalan dengan penelitian sebelumnya dimana estimasi dengan metode berbasis *maximum likelihood* yang umum digunakan dalam CFA atau SEM dapat menduga parameter-parameter dengan menghasilkan probabilitas dari data dengan hasil yang terbaik (Heck & Thomas, 2015) dimana jumlah sampel minimum yang digunakan pada penelitian ini sudah jauh memenuhi syarat yang ada.

## Penutup

Berdasarkan penjelasan mengenai hasil dalam penelitian ini, maka berdasarkan hal tersebut dapat disimpulkan sebagai berikut:

1. Berdasarkan hasil *first order* CFA pada alat ukur *social quality of life* terdapat 22 item yang fit mengukur *social quality of life*.
2. Berdasarkan pendeteksian *differential item functioning* (DIF) dengan pendekatan *multiple indicator multiple causes* (MIMIC) model, terdapat dua item yang mengandung DIF, yaitu item 5 dan item 23.
3. Bahwasanya *multiple indicator multiple causes* (MIMIC) model efektif dalam mendeteksi *differential item functioning* (DIF).

## Daftar Pustaka

- Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey: Brooks Cole.
- Anastasi, A. Dan Urbina, U., 1997, *Psychological testing* (seventh edition), Prentice- Hall Inc, New Jersey.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons, Inc.

- Brown, A Timothy. (2006). *Confirmatory factor analysis for research*. New York: The Guilford Press.
- Cai, L. (2010). Metropolis-hastings robbins-monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35, 3, 307-335.
- Cai, L. (2010b). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, 75, 33-57.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Carr, A. (2004). *Positive psychology: the science of happiness and human strengths*. New York, NY: Brunner-Routledge.
- Cheng, Y., Shao, C. & Lathrop, Q. N. (2015). The mediated MIMIC model for understanding the underlying mechanism of DIF. *Educational and Psychological Measurement*, 76, 1, 43-63.
- Crane, Gibbons, Narasimhalu, Lai & Cella. (2007). Rapid detection of differential item functioning in assessments of health-related quality of life: The functional assessment of cancer therapy. *Quality Life Research*, 16, 101-114.
- Felce, D., & Perry, J. (2005). Exploring current conceptions of quality of life: a model for people with and without disabilities. Dalam Renwick, I. Brown, & M. Nagler (Eds.), *Quality of life in health promotion and rehabilitation: conceptual approaches, issues, and applications*. California: SAGE Publication.
- Finch, Holmes. (2012). The MIMIC model as a method for detecting DIF: comparison with mantel-haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement*, 278-294
- Fox, Jean-Paul. (2010). *Bayesian item response modeling: Theory and applications*. New York: Springer.
- Gallo, J. J., Anthony, J. C., & Muthen, B. O. (1994). Age differences in the symptoms of depression: A latent trait analysis. *Journal of Gerontology: Psychological Sciences*, 49, 251-264.
- Heck, H. R., & Thomas, L. S. 2005. *An introduction to multilevel modeling techniques*. New York: Routledge Taylor & Prancis Group
- Kaplan, R. M. & Saccuzzo. (2005). *Psychological Testing: principles, application, and issues (6th ed.)*. Belmont : Thomson Wadsworth.
- Lee, S. Y. (2007). *Structural equation modeling: a bayesian approach*. Chicester: John Wiley & Sons Ltd.
- Muthen, B. O. 1989. Latent variable modeling in heterogenous populations. *Psychometrika*, 54, No. 4, 557-585
- Power, M. J. (20013). Quality of Life. dalam Lopez, S. J., & Synder, C. R. (Ed). *Positive psychological assessment: a handbook of models and measures*. Washington: American Psychological Association.
- Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life items banks: Plans for the patient-reported outcome measurement information system (PROMIS). *Medical Care*, 45, S22-S31.
- Sarafino, E. P. & Smith, T. W. (2011). *Health psychology: biopsychosocial interactions (7th edition)*, Hoboken, NJ: John Wiley & Sons, Inc.
- Scot, W Neil. Fayers, M Peter. Aaronson, K Neil. Bottomey, Andrew. Graeff, de Alexander. Groenvold, Mogens. Gundy, Chad. Koller, Michael. Pettersen, A Morten. & Sprangers, AG Mirjam. (2010). Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health and Quality of Life Outcomes*, 8, 1-9

- Sills, L.C., & Brown, T.A. (2006). Research considerations: latent variable approaches to studying the classification and psychopathology of mental disorders. In M. Hersen, J. C. Thomas & F. Andrasik (Eds.), *Comprehensive Handbook of Personality and Psychopathology*. Volume 2. Hoboken, N.J.: John Wiley & Sons.
- Simon, M. K. & Goes, J. (2013). *Dissertation and scholarly research: recipes for success*. Seattle, WA: Dissertation Success LLC.
- Taylor, S.E., Peplau, L.A., Sears, D.O. 2000. *Social psychology, 10<sup>th</sup> edition*. USA : Prentice Hall
- Wang, W. C., Shih, C. L., & Yang, C. C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement*, 69, 713-731.
- Wang, Jichuan. & Wang, Xiaoqian. (2012). *Structural equation modelling: application using Mplus*. United Kingdom: A John Wiley & Sons
- Umar, J. (2011). *Bahan ajar statistik*. Fakultas Psikologi UIN Jakarta. Tidak dipublikasikan.
- Woods, Oltmanns & Turkheimer. (2008). Illustration of MIMIC-Model DIF testing with the schedule for nonadaptive and adaptive personality. *Journal Psychopatologi Behaviour Assessment*, 31, 320-330.