

## PENGUJIAN KESETARAAN PRESISI DAN SKALA UKUR BUTIR-BUTIR PADA SKALA PSIKOLOGI

*Wahyu Widhiarso*

Fakultas Psikologi Universitas Gadjah Mada

Email: wahyupsy@gmail.com

### ABSTRACT

The purpose of this study is to identify the measurement models that appropriate to measure psychological attributes toward psychological scale. Measurement models that includes is parallel, tau-equivalent and congeneric models. Each model has different assumptions according to each item's precision and scale measuring on psychological scale. As a sample of psychological scale, this study uses Child Depression Inventory (CDI) and Rosenberg Self-Esteem Scale. Data was taken from the 3.183 adolescents from junior and senior high school in DIY, with ages ranging from 15 to 19. Both scale were tested using the Hotelling T2 test for testing the equality of mean and variance as well as confirmatory factory analysis through structural equation modeling to rest the accuracy of the model. This study found that psychological scales tend to fit better with the congeneric model than others. The congeneric model is one which assumes that every item in a scale has a different precision and scale of measure. Therefore, it is concluded that items tend to be measured in psychological attributes, with relative variation.

*Keywords: Measurement Model, Psychological Scale, Items Precision and Scale Measure*

### ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi model pengukuran yang tepat dengan pengukuran melalui skala psikologi. Model pengukuran yang dilibatkan adalah model pengukuran model paralel, nilai tau setara dan konjenerik. Masing-masing model memiliki asumsi yang berbeda terhadap presisi dan ukur skala butir-butir dalam satu skala. Skala psikologi yang dipakai sebagai sampel adalah Inventori Depresi Anak (CDI) dan Skala Harga Diri. Data diambil dari 3.183 remaja SMP dan SMA dengan rentang usia 15 hingga 19 tahun di DIY. Uji statistik yang dilibatkan adalah uji Hotelling T2 untuk menguji kesetaraan rerata dan varians serta analisis faktor konfirmatori melalui pemodelan persamaan struktural untuk menguji ketepatan model. Penelitian ini menemukan bahwa model konjenerik memiliki ketepatan model yang lebih tinggi dibanding dengan kedua model lainnya. Model konjenerik adalah model pengukuran yang mengasumsikan bahwa setiap butir dalam satu skala memiliki presisi dan skala ukur yang berbeda-beda. Dengan demikian dapat disimpulkan bahwa butir-butir dalam pengukuran melalui skala psikologi memiliki daya ukur yang relatif bervariasi.

**Kata Kunci: Model Pengukuran, Skala Psikologi, Presisi dan Skala Ukur Butir**

### PENGANTAR

Skala psikologi banyak dikembangkan dengan mengasumsikan bahwa setiap butir di dalam skala tersebut memiliki ketepatan ukur yang setara. Hal ini terlihat dari banyaknya pengembangan skala-skala psikologi

dievaluasi dengan menggunakan koefisien alpha sebagai ukuran keandalan pengukuran skala. Koefisien alpha dipakai untuk mengidentifikasi konsistensi internal butir-butir di dalam skala. Dengan menggunakan koefisien alpha sebagai acuannya, secara

tidak langsung para pengembang skala mengikuti apa yang diasumsikan oleh koefisien alpha yaitu butir-butir di dalam skala memiliki kesamaan presisi dan skala dalam mengukur. Dalam bahasa psikometri kesetaraan presisi ukur tersebut dinamakan dengan *tau* setara (*tau-equivalent*) (Salkind, 2010). *Tau* adalah huruf Yunani yang dalam psikometri menjelaskan skor murni (T).

Berbagai penelitian menemukan bahwa asumsi kesamaan atau kesetaraan kapasitas ukur tersebut tidak mudah untuk dipenuhi (e.g. Lucke, 2005). Jika asumsi ini tidak dipenuhi maka koefisien reliabilitas yang dihasilkan berada pada nilai di batas estimasi terendah (*underestimate*). Dalam butir pengukuran depresi melalui Inventori Depresi dari Kovac (1985) misalnya, butir yang menanyakan gejala bunuh diri dan butir yang menanyakan kurangnya nafsu makan memiliki presisi ukur yang berbeda dalam mengukur depresi. Butir pertama memiliki lebih target yang lebih memusat dibanding dengan butir kedua. Hal ini dapat menyebabkan rerata dan varians skor dari kedua butir ini berbeda. Contoh lainnya adalah pengukuran kualitas hidup melalui Skala SF-36 dari Ware (1993) yang sering dipakai dalam mengukur kualitas hidup. Butir yang menanyakan keberfungsian badaniah memiliki kapasitas ukur yang berbeda dengan keberfungsian sosial dalam mengukur kualitas hidup individu.

Ketidaksetaraan kapasitas ukur butir-butir dalam psikologi diakibatkan oleh sebagai berikut, *Pertama*, skala psikologi dikembangkan dari teori mengenai atribut yang hendak diukur yang kebanyakan tidak menyertakan secara mendetail operasionalisasi konsep teoritik di dalamnya menjadi indikator perilaku (DeVellis, 1991). Para pengembang skala sendiri yang banyak berperan dalam mengoperasionalkan konsep teoretik tersebut menjadi indikator perilaku yang nantinya akan diturunkan menjadi butir dalam skala psikologi. *Kedua*, tidak ada kaidah yang

baku dalam menurunkan indikator menjadi kalimat pernyataan (Colton & Covert., 2007). Dengan tidak adanya kaidah baku penulisan butir, maka perilaku memukul sebagai indikator perilaku agresi dapat diturunkan menjadi butir "*Saya akan memukul orang yang menghina saya*" atau butir "*Saya akan memukuli orang yang menghina saya*". Jelas sekali bahwa meski sama-sama memukul namun kedua butir memiliki presisi ukur perilaku agresi yang berbeda. *Ketiga*, setiap pernyataan di dalam butir memiliki kecenderungan yang berbeda-beda untuk disetujui oleh responden (Bradburn, 2004). Antara pernyataan yang berisi opini normatif, misalnya "*Remaja seharusnya....*" dan pengalaman "*Ketika pada masa remaja saya pernah...*", memiliki perbedaan kecenderungan yang berbeda untuk disetujui oleh responden. Penelitian Widhiarso dan Suhapti (2009) menemukan bahwa perbedaan sumber penilai antara sumber dari penilaian responden sendiri, misalnya "*Saya adalah....*", dan sumber dari orang lain, misalnya "*Menurut teman saya, saya adalah...*", turut mempengaruhi perbedaan respons dari responden. Perbedaan kecenderungan ini selain terkait dengan pernyataan di dalam butir, juga terkait dengan situasi pengukuran.

*Keempat*, kesetaraan pengukuran hanya dapat dicapai ketika pengukuran bersifat unidimensi (Green & Yang, 2009). Hasil penelitian menunjukkan bahwa pengukuran dalam bidang psikologi cenderung bersifat multidimensi (Brunner & Süß, 2005; Kamata, Turhan, & Darandari, 2003); ditambah lagi dengan koefisien alpha tidak sensitif terhadap kemajemukan dimensi (Sijtsma, 2009). Banyak pengukuran yang sebenarnya bersifat multidimensi namun tetap memiliki koefisien alpha yang tinggi. Ada beberapa penyebab mengapa pengukuran psikologi yang bersifat unidimensi sulit untuk dicapai ketika skala memiliki butir-butir yang banyak. Penyebab tersebut antara lain karakteristik alamiah konstruk psikologi, adanya pelibatan aspek-

aspek dalam penyusunan alat ukur, jumlah butir di dalam skala pengukuran, teknik penulisan butir dan satuan pengukuran yang berbeda (Widhiarso, 2009).

Kecenderungan bahwa butir-butir skala psikologi memiliki presisi dan skala ukur yang bervariasi perlu diverifikasi lebih lanjut. Teori psikometri terbaru telah menyediakan berbagai model yang dapat mengakomodasi kompleksitas teknik pengukuran dengan menggunakan skala psikologi. Tujuannya adalah agar properti psikometris skala psikologi yang dihasilkan benar-benar mewakili hasil pengukuran yang didapatkan.

### Model Pengukuran dalam Bidang Psikologi

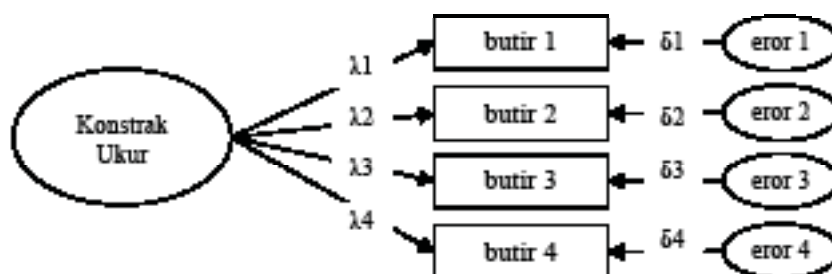
Dalam literatur psikologi telah dikenal empat model pengukuran yang masing-masing mengasumsikan setara tidaknya presisi butir pengukuran, kesamaan skala dan sesatan varians pengukuran. Keempat model tersebut adalah model pengukuran paralel, kesetaraan nilai tau, kesetaraan nilai tau esensial, dan konjenerik (Lucke, 2005). Kesamaan presisi atau ketepatan pengukuran ditunjukkan oleh kesamaan rerata skor sedangkan kesamaan skala ditunjukkan dengan kesamaan nilai varians (Raykov, 2001).

Model paralel adalah model yang paling ketat dengan mengasumsikan bahwa

skala dan presisi pengukuran oleh butir serta varians sesatan pengukuran (*error measurement*) adalah sama. Model kesetaraan nilai tau (*tau-equivalent*) mengasumsikan bahwa skala dan presisi pengukuran oleh ukur butir adalah sama akan tetapi varians sesatannya boleh berbeda. Model kesetaraan nilai tau esensial (*essentially tau-equivalent*) mengasumsikan bahwa pengukuran oleh item harus dalam skala sama, akan tetapi memiliki presisi dan varians sesatan yang berbeda. Model konjenerik (*congeneric*) memiliki asumsi yang lebih moderat karena memperbolehkan skala, presisi pengukuran dan varians sesatan pengukuran boleh berbeda (Graham, 2006).

Dengan demikian untuk memenuhi asumsi paralel, skor butir harus memiliki rerata dan varians skor butir yang sama. Asumsi ini sulit untuk dipenuhi karena rerata dan varians yang sama biasanya didapatkan dari butir yang memiliki target indikator perilaku yang sama, padahal skala pengukuran mewakili dari domain ukur yang memiliki perilaku yang majemuk. Memfokuskan pada sedikit indikator saja akan menyebabkan pengukuran menjadi tidak komprehensif. Salah satu cara untuk mengatasi hal ini adalah penggunaan model yang tidak terlalu ketat dalam mengasumsikan skor hasil pengukuran.

Model pengukuran banyak dikaji dalam pemodelan persamaan struktural (SEM)



Gambar 1.  
Model Pengukuran

karena model pengukuran merupakan salah satu bagian dari elemen SEM selain model struktural. Gambar 1 menunjukkan ada empat butir yang mengukur satu faktor konstrak ukur. Di dalam skor tiap butir di dalamnya terkandung dua komponen, yaitu komponen dari konstrak yang diukur yang disimbolkan dengan  $\lambda_i$  dan komponen sesatan pengukuran, yang disimbolkan dengan  $\delta_i$ . Model ini didasari dari teori klasik psikometri yang menyatakan bahwa di dalam varians skor tampak terkandung varians skor murni dan varians sesatan pengukuran (Raykov & Mels, 2009).

Model pengukuran dalam pendekatan SEM diformulasikan dalam analisis faktor konfirmatori (CFA) yang berisi bobot tiap butir pada faktor ( $\lambda_i$ ) dan varians yang tidak terkait dengan faktor ( $\delta_i$ ). Dengan demikian, dari model pengukuran di atas dapat diketahui bahwa model paralel yang mengasumsikan bahwa tiap butir mengukur konstrak sama, memiliki presisi skala, dan sesatan pengukuran sama. Model paralel ditunjukkan dengan persamaan yang menunjukkan kesamaan bobot ukur tiap butir :  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$  dan kesamaan varians eror pengukuran  $\delta_1 = \delta_2 = \delta_3 = \delta_4$ . Model tau setara dan konjenerik juga dapat diformulasikan melalui persamaan tersebut.

Penelitian ini bertujuan untuk menguji penerapan tiga model pengukuran, yaitu model paralel, kesetaraan nilai tau, dan konjenerik melalui analisis faktor konfirmatori. Dalam pendekatan SEM, ketepatan penerapan model tersebut ditunjukkan melalui indeks ketepatan model (*goodness fit indices*). Dari paparan teoritik yang menunjukkan bahwa butir-butir dalam skala psikologi cenderung memiliki presisi ukur, skala ukur, dan sesatan yang bervariasi. Maka dari itu, peneliti menghipotesiskan bahwa model pengukuran yang moderat (i.e konjenerik) lebih menggambarkan hasil pengukuran psikologi dibanding dengan model pengukuran yang ketat (i.e paralel). Indeks ketepatan model

pada model konjenerik akan lebih tinggi dibanding dengan indeks ketepatan pada model paralel.

## Partisipan

*Remaja.* Data diambil dari penelitian yang dilakukan oleh Retnowati (2003) yang melibatkan remaja sekolah antara usia 15 tahun hingga 19 tahun. Rata-rata usia partisipan adalah 16 tahun dengan jumlah 3.183 orang yang terdiri dari 1.474 (46%) pria dan 1.709 (54%) wanita. Partisipan adalah siswa SLTP, SMU, SMK dan PSBR di Daerah Istimewa Yogyakarta dari empat wilayah antara lain Kabupaten Sleman, Bantul, Kulon Progo, Gunung Kidul dan Kodya Yogyakarta. Masing-masing Kabupaten dan Kodya Yogyakarta, diwakili oleh beberapa SMU dan SMK.

## Pengukuran

Inventori Depresi Anak diadaptasi dari *Children Depression Inventory* (CDI) yang merupakan inventori untuk mengungkap simptom depresi pada anak dan remaja atau dengan rentang usia 7 sampai 19 tahun, yang meliputi kesedihan, *anhedonia*, ide bunuh diri, dan gangguan nafsu makan. CDI terdiri dari 27 butir dapat digunakan untuk anak berusia sekitar 7 sampai dengan 19/20 tahun. Di samping skala dalam bentuk panjang, juga tersedia skala dalam bentuk pendek yang terdiri atas 12 butir (Carlson & Cantwell, dalam Matson, 1989). Uji coba CDI pertama kali dilakukan penulis, dengan menggunakan kriteria eksternal dan kriteria internal. Uji coba dilakuka pada 109 subjek, dengan kriteria eksternal, yaitu dengan cara mengkorelasikan dengan BDI (*Beck Depression Inventory*) dengan hasil  $r=0.561$  ( $p<0.01$ ). Uji coba dilakukan lagi pada subjek sebanyak 252 orang, dengan kriteria internal, hasil menunjukkan indeks daya beda yang berkisar antara 0.1721 sampai dengan 0.3795, dengan koefisien reliabilitas konsistensi intenal sebesar 0.7135.

*Skala Harga Diri.* Alat ukur untuk mengungkap harga diri (*self esteem*) yang digunakan dalam penelitian ini dikembangkan oleh Rosenberg (1965). Rosenberg mengoperasionalkan konsepnya dalam bentuk 10 butir. Responden diminta untuk memberi jawaban berdasar kriteria Guttman yaitu, sangat setuju; setuju; tidak setuju dan sangat tidak setuju. Pernyataannya antara lain: “Secara keseluruhan saya puas dengan diri saya”; “Saya pikir saya sama sekali tidak baik”. Hasil analisis butir skala harga diri menunjukkan dari 10 butir yang diujicobakan, korelasi butir total berkisar antara 0,2581 – 0,3917 dengan koefisien reliabilitas sebesar 0,8689.

### Analisis Data

Penelitian ini terdiri dari tiga tahap yaitu menguji kesetaraan rerata dan varians butir pada skala yang sama, mengidentifikasi ketepatan tiga model pengukuran pada tiap skala dan menguji perbedaan ketepatan antarketepatan model. Pengujian kesetaraan rerata skor butir dilakukan dengan menggunakan uji Hotelling T<sup>2</sup> sedangkan pengujian kesetaraan varians skor butir dilakukan dengan menggunakan uji kai-kuadrat. Kedua uji ini dilakukan dengan menggunakan program bantu SPSS versi 16 (SPSS Inc, 2007).

Identifikasi ketepatan model dilakukan dengan menggunakan analisis faktor konfirmatori. Tiga model pengukuran yaitu model paralel, nilai tau setara dan konjenerik diterapkan pada tiap skala pengukuran. Tiga indeks ketepatan model yang dipakai adalah kai-kuadrat, *goodness fit index* (GFI), dan *Root Means Square Error* (RMSEA). Analisis hanya diarahkan pada identifikasi ketepatan model saja karena penelitian ini tidak untuk mengembangkan model pengukuran. Identifikasi model pengukuran dilakukan dengan menggunakan analisis faktor konfirmatori melalui teknik estimasi kebolehdian maksimal (*maximum likelihood*) yang dilakukan dengan program analisis

LISREL 8.30. Indeks ketepatan model yang didapatkan akan dibandingkan untuk mendapatkan apakah ada perbedaan ketepatan model yang signifikan antara satu model pengukuran dengan model pengukuran lainnya.

## PEMBAHASAN

### Uji Kesetaraan Rerata dan Varians

Uji kesetaraan rerata dilakukan dengan menggunakan uji Hotelling T<sup>2</sup>, dengan hipotesis nihil bahwa semua butir dalam satu skala/faktor skala memiliki rerata yang sama. Uji kesetaraan varians dilakukan dengan menggunakan uji kai-kuadrat dengan hipotesis nihil bahwa semua butir dalam skala memiliki varians dan varians sesatan yang setara.

Hasil analisis yang dipaparkan pada Tabel 1 menunjukkan bahwa dalam satu skala atau faktor dalam skala, cenderung memiliki rerata, varians butir, dan varians sesatan yang berbeda-beda. Uji kesetaraan rerata menghasilkan bahwa pada semua butir dalam skala pengukuran memiliki varians dan varians sesatan yang berbeda. Nilai kai-kuadrat ( $\chi^2$ ) dan nilai F yang dihasilkan cukup besar sehingga hipotesis nihil yang mengatakan bahwa rerata, varians dan varians sesatan, ditolak.

Temuan penelitian ini mendukung asumsi bahwa skala psikologi cenderung memiliki presisi pengukuran, skala

Tabel 1.  
Hasil Pengujian Kesetaraan Rerata, Varians dan Varians Sesatan Butir

Skala Pengukuran	Uji Kesetaraan Rerata		Uji Kesetaraan Varians	
	T2	db	$\chi^2$	db
BDI Faktor 1	1057.355		675.409	19
BDI Faktor 2	2909.784	3	1215.921	8
BDI Faktor 3	47.876	3	527.742	8
BDI Faktor 4	412.256	7	1083.693	34
BDI Faktor 5	1868.608	4	538.239	13
Roosenberg	1253.587	9	3070.782	53

Keterangan : Semua uji perbandingan menunjukkan hasil signifikan pada taraf 1% ( $p < 0.01$ )

pengukuran, dan sesatan pengukuran yang berbeda-beda. Namun demikian hasil ini belum menjawab model pengukuran mana yang tepat dalam menggambarkan data hasil pengukuran. Oleh karena itu, penulis melanjutkan analisis pada pengujian ketepatan model pada masing-masing model pengukuran.

### Perbandingan Ketepatan Model Pengukuran

*Pengukuran Depresi.* Pengujian ketepatan model dalam pengukuran depresi dibagi menjadi lima sesuai dengan faktor-faktor depresi. Hasil analisis dipaparkan pada Tabel 2 yang menunjukkan perbandingan model pengukuran pada tiap faktor pengukuran depresi. Pada semua faktor depresi, model pengukuran konjenerik memiliki ketepatan pengukuran yang lebih tinggi dibanding dengan model paralel, dan kesetaraan nilai tau. Perubahan model dari paralel, nilai tau setara hingga konjenerik menunjukkan peningkatan nilai GFI, sebaliknya nilai RMSEA menunjukkan penurunan. Hal ini menunjukkan bahwa perubahan model tersebut meningkatkan ketepatan model dengan data.

Dari semua faktor pengukuran depresi pada model pengukuran konjenerik, ada dua faktor yang kesemua indeks ketepatannya

di atas batas penerimaan model ( $GFI > 0.9$  &  $RMSEA < 0.08$ ), sebaliknya ada tiga faktor yang masih ada indeks ketepatan model di bawah yang direkomendasikan. Namun demikian hasil ini tidak di bahas dalam penelitian ini karena fokus pembahasan pada perbandingan ketepatan model yang dilakukan pada analisis berikut.

Tabel 3 menunjukkan hasil perbandingan ketepatan model dengan menggunakan uji kai kuadrat. Semua hasil analisis menunjukkan adanya perbedaan yang signifikan ( $p < 0.01$ ) pada semua faktor pengukuran depresi. Model nilai tau setara memiliki ketepatan model lebih tinggi dibanding model paralel, dan model konjenerik memiliki ketepatan model lebih tinggi dibanding model paralel dan nilai tau setara. Dengan demikian dapat disimpulkan bahwa model konjenerik menggambarkan data lebih baik dibanding dengan model *paralel* dan atau setara.

Pada pengukuran harga diri didapatkan hasil yang sama dengan pengukuran depresi. Tabel 4 menunjukkan berubahnya model dari paralel, nilai tau setara hingga konjenerik diikuti dengan meningkatnya ketepatan model. Nilai GFI, AGFI dan CFI meningkat, sebaliknya nilai RMSEA menurun yang menunjukkan model yang dimodifikasi semakin mendekati data.

Tabel 2.  
Perbandingan Ketepatan Model Pengukuran pada Tiap Faktor CDI

Ketepatan Model	Faktor 1			Faktor 2			Faktor 3		
	Par	Tau	Kon	Par	Tau	Kon	Par	Tau	Kon
$\chi^2$	675.93	553.60	55.15	1216.54	493.56	33.07	528.01	820.86	153.10
db	19	14	9	8	5	2	8	5	2
GFI	0.93	0.94	0.99	0.93	0.93	0.93	0.93	0.89	0.97
RMSEA	0.11	0.12	0.04	0.11	0.11	0.11	0.14	0.23	0.16

Ketepatan Model	Faktor 4			Faktor 5		
	Par	Tau	Kon	Par	Tau	Kon
$\chi^2$	675.93	553.60	55.15	1189.83	493.56	33.07
db	19	14	9	8	5	2
GFI	0.93	0.94	0.99	0.83	0.93	0.99
RMSEA	0.11	0.12	0.04	0.22	0.17	0.07

Keterangan : Par = Model Paralel; Tau = Model Nilai Tau Setara; Kon = Model Konjenerik

Tabel 3 (a)  
Uji Statistik Perbandingan Ketepatan Model Tiap Faktor CDI

Perbandingan	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5
Par vs Tau	122.32 (5)	722.98 (3)	292.85 (3)	122.33 (5)	696.27 (3)
Par vs Kon	620.78 (10)	1183.47 (3)	374.91 (6)	620.78 (10)	1156.76 (6)
Tau vs Kon	498.45 (5)	460.49 (2)	667.76 (3)	498.45 (5)	460.49 (3)

Keterangan : Par = Model Paralel; Tau = Model Nilai Tau Setara; Kon = Model Konjenerik. Semua uji perbandingan menunjukkan hasil signifikan pada taraf 1% ( $p < 0.01$ )

Hasil uji statistik perbandingan antar ketepatan model menunjukkan adanya perbedaan yang signifikan antar ketiga model pengukuran. Model pengukuran konjenerik memiliki ketepatan model secara signifikan dibanding dengan dengan model paralel dan tau setara. Penelitian ini bertujuan untuk membandingkan ketepatan tiga model pengukuran pada skala psikologi. Ketiga model tersebut adalah model paralel, tau setara dan konjenerik. Pengukuran yang dipakai sebagai sampel adalah pengukuran depresi (CDI) dan pengukuran harga diri. Penelitian ini menemukan bahwa model konjenerik memiliki ketepatan model yang lebih tinggi dibanding dengan kedua model lainnya. Dengan demikian model konjenerik lebih tepat dikenakan pada data hasil pengukuran psikologi.

Tabel 4.  
Perbandingan Ketepatan Model Pengukuran pada Pengukuran Harga Diri

Ketepatan Model	Paralel	Tau Setara	Konjenerik
Chi Square (p)	1144.39	793.70	602.14
GFI	0.76	0.81	0.85
db	53	44	35
RMSEA	0.17	0.16	0.15

Tabel 3 (b)  
Uji Statistik Perbandingan Ketepatan Model Pengukuran pada Tiap Faktor CDI

Perbandingan	Kai-Kuadrat
Paralel vs Kesetaraan Nilai Tau	122.32 (5)
Paralel vs Kojenerik	620.78 (10)
Konjenerik vs Kesetaraan Nilai Tau	498.45 (5)

Keterangan : Semua uji perbandingan menunjukkan hasil signifikan pada taraf 1% ( $p < 0.01$ )

Model konjenerik memiliki asumsi yang lebih moderat dibanding dengan model paralel dan nilai tau setara. Model konjenerik tidak mensyaratkan rerata, varians dan varians sesatan yang setara antara satu butir dengan butir lainnya. Model konjenerik mampu mengakomodasi karakteristik skala psikologi yang memiliki keunikan antara satu butir dengan butir lainnya. Dalam hal ini keunikan tersebut adalah presisi dan skala ukur butir dalam skala yang cenderung bervariasi. Hasil penelitian ini didukung oleh beberapa pernyataan peneliti yang mengkaji mengenai model konjenerik (Graham, 2006; Raykov, 2001).

Verifikasi asumsi kesetaraan rerata dan varians skor butir yang dilakukan dalam penelitian ini menunjukkan bahwa skala psikologi cenderung memiliki presisi dan skala ukur yang berbeda. Hasil ini diperkuat dengan hasil analisis faktor konfirmatori yang menunjukkan bahwa setiap butir memiliki bobot faktor yang bervariasi antara satu butir dengan butir lainnya. Bobot faktor yang berbeda-beda sekaligus menunjukkan bahwa setiap butir memiliki sumbangan efektif yang berbeda-beda dalam menjelaskan atribut ukur. Penelitian ini membuktikan bahwa butir yang mengukur perilaku yang umum memiliki sumbangan efektif yang besar dibanding butir yang memusat pada perilaku yang khusus. Misalnya pada butir CDI, butir yang mengukur kesedihan secara umum, memiliki bobot faktor yang lebih besar dibanding dengan butir yang mengukur seberapa jauh peristiwa buruk disebabkan oleh kesalahan subjek. Hal ini terjadi karena butir yang pertama memiliki cakupan yang

luas sehingga memiliki variasi yang besar dibanding dengan butir kedua. Hasil analisis faktor konfirmatori menunjukkan bahwa model yang mengasumsikan bahwa setiap butir memiliki keunikan dalam hal presisi dan skala ukur memiliki nilai ketepatan model yang lebih tinggi. Setiap butir yang diperkenankan untuk memiliki nilai bobot faktor dan sesatan pengukuran yang berbeda-beda lebih tepat dibanding dengan memaksakan bahwa butir-butir memiliki presisi dan skala ukur yang sama.

Program bantu analisis seperti SPSS (2007) telah memfasilitasi upaya peneliti untuk memverifikasi asumsi data hasil pengukuran. Menu model paralel memverifikasi apakah butir memiliki kesetaraan dalam hal varians dan varians sesatan, sedangkan menu model paralel ketat (*strict paralel*) memverifikasi kedua asumsi dalam model paralel ditambah dengan asumsi kesetaraan nilai rerata. Verifikasi kesetaraan rerata dan varians skor butir tersebut dilaporkan pada bagian ketepatan (*goodness fit index*) melalui nilai kai-kuadrat. Nilai kai-kuadrat yang signifikan menunjukkan bahwa ada perbedaan yang signifikan antar rerata dan varians skor antarbutir yang dianalisis. Selain itu ada uji Hotelling  $T^2$  yang dapat dipakai untuk menguji kesetaraan rerata saja.

Hasil penelitian ini merekomendasikan bahwa peneliti perlu memverifikasi data hasil pengukuran sebelum mengidentifikasi properti psikometris berdasarkan formula yang tepat dengan karakteristik data dan tidak bergantung pada koefisien tertentu secara monoton. Pernyataan ini didukung oleh Ferketich (1990) yang mengatakan bahwa seharusnya kajian dan pengujian reliabilitas tidak hanya terpaku pada satu koefisien saja melainkan juga melibatkan koefisien lain yang kemungkinan menggambarkan hasil yang lebih optimal. Socan (2000) mengatakan bahwa banyak di antara para peneliti yang hanya terpaku pada penggunaan koefisien

Alpha Cronbach dalam mengestimasi reliabilitas dengan menggunakannya secara monoton tanpa memperhatikan asumsi yang melatarbelakanginya.

Pemilihan formula estimasi reliabilitas secara monoton tersebut dapat diakibatkan oleh dua sebab, pertama minimnya pemahaman peneliti mengenai koefisien reliabilitas yang dapat menjadi alternatif, kedua, minimnya keberadaan program komputasi yang dapat mengelaborasi model pengukuran yang mereka susun dengan mudah. Feldt (1987) mengatakan bahwa popularitas koefisien alpha lahir karena beberapa faktor, antara lain: a) teknik komputasinya relatif mudah karena hanya memerlukan informasi berupa varian butir dan varian skor total, b) distribusi sampling sudah diketahui sehingga penentuan interval kepercayaan pada populasi sangat dimungkinkan. Koefisien alpha juga banyak dipakai pada banyak literatur karena merupakan estimator yang moderat dalam mengestimasi reliabilitas.

Pengembangan skala psikologi dengan teori klasik menggunakan pendekatan analisis dengan asumsi yang lebih moderat dan dapat diakomodasi oleh analisis pemodelan persamaan struktural (SEM) (Albright, 2006; Raykov, 2009). Melalui pendekatan ini peneliti memiliki keleluasaan untuk mengembangkan model yang sesuai dengan data yang dimilikinya. Ketika skala semua butir terbukti memiliki presisi dan skala ukur yang sama maka peneliti dapat melakukan pembatasan (*constraint*) parameter butir di dalam model, demikian juga ketika mendapati butir di dalam skala memiliki presisi dan skala ukur yang berbeda. Meskipun asumsi-asumsi statistik berada dalam tataran teoritik yang bagi sebagian pakar tidak perlu diverifikasi ketika diterapkan pada tataran praktis (e.g. Azwar, 2000), namun verifikasi data akan memberikan hasil estimasi yang tepat sehingga perlu untuk dilakukan.



## SIMPULAN

Hasil penelitian ini menunjukkan bahwa butir-butir dalam skala psikologi lebih cenderung memiliki presisi ukur yang tidak setara. Secara teknis dalam perspektif pemodelan persamaan struktural, perbedaan presisi ini terlihat dari bobot faktor dan varians eror yang tidak setara. Saran yang dapat diberikan kepada peneliti adalah agar melakukan prosedur estimasi reliabilitas setelah mengidentifikasi model pengukuran terlebih dahulu. Sesuai dengan prosedur baku penyusunan skala (e.g Netemeyer, Bearden, & Sharma, 2003; Spector, 1992), analisis faktor baik eksploratori maupun konfirmatori dilakukan terlebih dahulu sebelum reliabilitas pengukuran diestimasi. Analisis faktor selain bertujuan untuk melihat dimensionalitas pengukuran (unidimensional vs. multidimensinal) juga mengeksplorasi ragam presisi dan skala ukur tiap butir yang terlihat melalui bobot faktor tiap butir.

## DAFTAR PUSTAKA

- Albright, J, 2006, *Confirmatory Factor Analysis using Amos, Lisrel, and MPLUS*, The Trustees of Indiana University.
- Azwar, S, 2000, *Asumsi-asumsi dalam inferensi statistika*, Manuskrip tidak dipublikasikan, Yogyakarta: Fakultas Psikologi UGM.
- Bradburn, N. M, 2004, *Asking questions : the definitive guide to questionnaire design*, San Francisco, CA: John Wiley & Sons, Inc.
- Brunner, M., & Süß, H. M, 2005, *Analyzing the reliability of multidimensional measures: An example from intelligence research*, *Educational and Psychological Measurement*, 65(2), hlm.227-240.
- Colton, D., & Covert., R. W, 2007, *Designing and constructing instruments for social research and evaluation*, San Francisco, CA: John Wiley & Sons, Inc.
- DeVellis, R. F, 1991, *Scale development: Theory and applications*, Newbury Park: SAGE Publications, Inc.
- Ferketich, S, 1990, *Focus on Psychometrics Internal Consistency Estimates of Reliability*, *Researching Nursing & Health*, 13, hlm.437-440.
- Graham, J. M, 2006, *Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability*, *Educational and Psychological Measurement*, 66(6), hlm.930-944.
- Green, S., & Yang, Y, 2009, *Commentary on Coefficient Alpha: A Cautionary Tale*. *Psychometrika*, 74(1), hlm.121-135.
- Kamata, A., Turhan, A., & Darandari, E, 2003, *Estimating reliability for multidimensional composite scale ccores* Paper presented at the American Educational Research Association, Chicago, April 2003.
- Kovacs, M, 1985, *The Children's Depression Inventory (CDI)*, *Psychopharmacology Bulletin*, 21(4), hlm.995-998.
- Lucke, J. F, 2005, *The  $\alpha$  and the  $\omega$  of Congeneric Test Theory: An Extension of Reliability and Internal Consistency to Heterogeneous Tests*, *Applied Psychological Measurement*, 29(1), hlm.65-81.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S, 2003, *Scaling Procedures: Issues and Applications*, Thousand Oaks, CA: Sage Publications.
- Raykov, T, 2001, *Bias of Coefficient afor Fixed Congeneric Measures with Correlated Errors*, *Applied Psychological Measurement*, 25(1), hlm. 69-76.
- Raykov, T, 2009, *Evaluation of Scale Reliability for Unidimensional Measures Using Latent Variable Modeling*, *By: Raykov*, 42, hlm.223-232.
- Raykov, T., & Mels, G, 2009, *Interval Estimation of Interitem and Item-Total Correlations for Multiple Component Measuring Instruments With Ordinal Items*, *Structural Equation Modeling: A Multidisciplinary Journal*, 16, hlm.99-108.
- Salkind, N. J, 2010, *Encyclopedia of Research Design*, Thousand Oaks, CA: Sage Publications Inc.

- Sijtsma, K, 2009, Reliability Beyond Theory and Into Practice, *Psychometrika*, 74(1), hlm.169-173.
- Socan, G, 2000, Assessment of Reliability when Test Items are not Essentially t-Equivalent, In A. Ferligoj & A. Mrvar (Eds.), *Developments in Survey Methodology* Ljubljana: FDV.
- Spector, P. E, 1992, *Summated rating scaling construction: An introduction*, Newbury Park: Sage Publication.
- SPSS Inc, 2007, *SPSS Base 16.0 User's Guide*, Chicago, IL: SPSS Inc.
- Ware, J. E., Snow, K. K., Kolinski, M., & Gandek, B, 1993, *SF-36 Health survey manual and interpretation guide*, Boston, MA.: The Health Institute New England Medical Centre.
- Widhiarso, W. (2009). Koefisien reliabilitas pada pengukuran kepribadian yang bersifat multidimensi. *Psikobuana*, 1(1), 39 - 48.
- Widhiarso, W., & Suhapti, R, 2009, Eksplorasi karakteristik item skala psikologis yang rentan terhadap tipuan respon, *Jurnal Psikologi*, 36(1), hlm.73-91.