

PERAN PENGUKURAN DAN ANALISIS STATISTIKA DALAM PENELITIAN PSIKOLOGI

Jahja Umar

Institut Asesmen Indonesia

umarindo@me.com

Abstract

At least, there are four source of error which cause research result invalidity, especially in data analysis using statistics, those are sampling error, measurement error, specification error, other sources of error that haven't been know and assumed as random error. In psychology research, impact of measurement error to research result validity more dominantly is caused by almost all variables in psychology that are latent. Impact of measurement error generally is in form of under-estimation to statistical coefficient such as correlation, regression, etc. But now, there are methods that can be used to correct all those negative impacts, such as confirmatory factor analysis which result true-score estimation to be used as data that will be analysed. There is also structural equation model analysis where error impact can corrected directly, thus pure correlation and regression coefficient obtained. This article illustrates how serious error that can happen without correction, using SEM approach

Keywords: *Measurement, Measurement Error, Latent Variable, Regression, Structural Equation Modeling*

Abstrak

Setidaknya ada empat sumber kesalahan yang dapat mengakibatkan tidak validnya hasil penelitian, terutama dalam analisis data statistika, yaitu: (a) kesalahan sampling, (b) kesalahan pengukuran, (c) kesalahan spesifikasi/teori, dan (d) sumber kesalahan lain yang belum diketahui dan dianggap bersifat random. Dalam penelitian psikologi dampak kesalahan pengukuran terhadap validitas hasil penelitian lebih dominan karena hampir semua variabelnya merupakan variabel laten. Dampak dari kesalahan pengukuran umumnya berupa under-estimasi terhadap koefisien statistik seperti korelasi, regresi, dan lain-lainnya. Namun pada saat ini, telah tersedia berbagai metode statistika yang dapat digunakan untuk mengoreksi dampak negatif tersebut, misalnya analisis faktor konfirmatorik yang dapat menghasilkan estimate true-score untuk digunakan sebagai data yang akan dianalisis. Selanjutnya juga terdapat analisis model persamaan struktural dimana dampak kesalahan pengukuran dapat dikoreksi secara langsung sehingga diperoleh koefisien korelasi dan regresi yang murni. Dalam tulisan ini disajikan ilustrasi (dengan pendekatan SEM) betapa seriusnya kesalahan yang dapat terjadi jika tidak dilakukan koreksi.

Kata Kunci: *Pengukuran, Kesalahan Pengukuran, Variabel Laten, Regresi, Structural Equation Modeling*

Diterima: 11 Juli 2014

Direvisi: 23 Agustus 2014

Disetujui: 30 Agustus 2014

PENDAHULUAN

Pembahasan mengenai pengukuran tak dapat dipisahkan dari konsep tentang variabel, yaitu sesuatu yang bervariasi dari satu kasus ke kasus lainnya. Sebagai lawan kata dari variabel ialah konstan, yaitu sesuatu yang selalu sama atau tidak bervariasi dari satu kasus ke kasus lainnya. Yang layak untuk diteliti atau dianalisis datanya adalah variabel, bukan konstan. Sebagai contoh, upaya untuk mendeskripsikan dan menjelaskan tentang bagaimana dan mengapa sesuatu bervariasi, adalah selalu menjadi pokok bahasan di dalam ilmu tentang metode analisis data seperti statistika. Seperti telah banyak diketahui, statistika adalah ilmu tentang distribusi frekuensi dari satu atau beberapa variabel. Agar dapat dianalisis dengan statistika, data mengenai variabel harus dikumpulkan dalam jumlah banyak, baik dalam arti kasusnya yang banyak atau kejadiannya yang berulang-ulang pada satu kasus.

Pada dasarnya sesuatu dapat bervariasi hanya dengan dua cara, yaitu (1) bervariasi menurut jenis atau kategorisasinya, atau (2) bervariasi menurut besaran (*magnitude*)-nya. Sesuatu yang bervariasi menurut jenisnya dinamakan variabel kategorik atau variabel jenis. Kadang-kadang disebut pula sebagai variabel nominal. Variabel seperti jenis kelamin, latar belakang etnis seseorang, agama yang dianut, jenis pekerjaan, jenis warna, dan sebagainya, adalah contoh dari variabel kategorik. Pengumpulan data untuk variabel seperti ini tidak melibatkan proses pengukuran. Yang dilakukan hanyalah penghitungan (*counting*) atas frekuensi/banyaknya kejadian atau kasus pada setiap kategori yang ditetapkan. Data yang dihasilkan berupa angka-angka dalam bentuk bilangan bulat positif (*integer*). Tidak pernah dihasilkan data dengan angka pecahan (*fraction*) ataupun desimal. Juga tidak ada data dalam bentuk bilangan negatif. Yang diperoleh hanyalah bilangan bulat positif mulai dari 0, 1, 2, 3, dan seterusnya. Di sini, kualitas data ditentukan oleh tiga hal yaitu: (1) kejelasan dari definisi tentang masing-masing kategori, (2) setiap kategori harus bersifat saling *mutual exclusive* yang artinya tidak boleh suatu kasus/kejadian termasuk ke dalam lebih dari satu kategori, dan (3) kecermatan dalam memasukkan suatu kasus ke dalam kategori yang sesuai dengan definisinya.

Salah satu problem yang sering dihadapi dalam pengumpulan data variabel kategorik adalah ditemuinya kasus yang tak dapat dimasukkan ke dalam kategori yang tersedia, meskipun definisinya telah jelas. Sebagai ilustrasi, misalkan pada variabel kelompok etnis/suku bangsa. Seorang subyek memiliki ibu yang berdarah campuran antara suku Bugis dan Jawa, serta memiliki bapak yang berdarah campuran Batak dan Lampung. Termasuk kategori suku/ etnis manakah subyek tersebut? Cara yang lazim dilakukan untuk mengatasi masalah seperti ini ialah dengan membuat kategori tambahan yang diberi label "lain-lain" (*others*).

Problem lain yang kadang terjadi ialah kesalahan karena dimasukkannya suatu kasus ke dalam lebih dari satu kategori (*multiple counting*). Kesalahan seperti ini akan sangat mempengaruhi hasil analisis data.

Jika ini terjadi, ibaratnya menghitung jumlah kaki kucing di mana kaki depan ada dua, kaki belakang ada dua, kaki kiri ada dua, dan kaki kanan ada dua, sehingga totalnya ada delapan. Tentu saja analisis terhadap data seperti ini akan sangat menyesatkan.

Cara mengatasi masalah ini juga cukup sederhana, yaitu dengan mencocokkan jumlah total kasus/kejadian sebelum dan sesudah klasifikasi dilakukan. Kalau tidak sama, berarti telah terjadi kesalahan dalam melakukan klasifikasi. Jika jumlah kasus setelah klasifikasi lebih besar, berarti telah terjadi *multiple counting*. Sedangkan jika jumlahnya lebih kecil, berarti ada kasus yang belum terhitung.

Pada tingkatan statistika dasar, analisis yang dapat dilakukan terhadap data kategorik amatlah terbatas, karena seperti telah dikemukakan sebelumnya, angka yang diperoleh hanyalah bilangan bulat positif (*integer*) dan bukanlah bilangan riil. Biasanya terbatas pada grafik atau tabel berdasarkan frekuensi atau proporsi/persentase. Namun pada tingkatan statistika lanjut, sekarang sudah cukup banyak tersedia metode analisis untuk melakukan prediksi maupun eksplanasi terhadap variabel kategorik, terutama melalui model yang bersifat probabilistik serta melalui berbagai bentuk transformasi atas data yang berbentuk proporsi tersebut. Selanjutnya adalah pembahasan mengenai sesuatu yang bervariasi menurut besaran (*magnitude*), yang disebut sebagai variabel kontinum. Dinamakan demikian karena suatu besaran biasanya digambarkan dalam bentuk kontinum pada suatu garis lurus. Artinya, satu ujung garis tersebut menunjukkan besaran dengan angka yang tinggi dan ujung lainnya menunjukkan besaran dengan angka yang rendah. Garis tersebut merupakan garis bilangan riil, dan kasus-kasus bervariasi dalam hal kedudukan atau posisi mereka pada garis kontinum tersebut. Pada garis kontinum ini dapat ditetapkan titik nol dan dapat dipilih suatu satuan ukuran (unit) untuk mengukur jarak suatu lokasi (pada garis) dari titik nol tersebut. Oleh sebab itu, dalam konteks pengumpulan data variabel kontinum, garis bilangan ini disebut skala pengukuran.

Kegiatan pengumpulan data variabel kontinum dilakukan dengan cara menentukan di mana posisi suatu kasus pada skala tersebut, dan kegiatan ini disebut pengukuran. Jadi dapat didefinisikan bahwa *pengukuran adalah kegiatan menetapkan posisi atau lokasi dari suatu obyek atau kasus pada suatu skala dengan menentukan berapa unit jaraknya dari titik nol pada skala tersebut*. Dari kegiatan ini selalu dihasilkan angka dalam bentuk bilangan riil karena suatu skala adalah garis bilangan riil. Angka yang lebih besar menunjukkan besaran yang lebih tinggi pada aspek yang diukur, dan sebaliknya, sesuai dengan definisi dari skala yang bersangkutan. Hal yang perlu diingat di sini ialah bahwa suatu skala harus memiliki titik nol (*scale origin*) dan satuan ukuran (*scaling unit*). Jika kedua hal ini belum ditetapkan, maka belum dapat dilakukan pengukuran.

Ditinjau dari segi titik nol yang ditetapkan, ada dua jenis skala ukuran yaitu: (1) skala rasio dan (2) skala interval. Suatu skala disebut skala rasio jika titik nol yang ditetapkan benar-benar menunjukkan tidak adanya kuantitas (besaran) dari aspek yang diukur. Titik nol seperti ini disebut nol absolut. Misalnya ketika

mengukur panjang meja dimulai dari salah satu ujungnya, atau mengukur tinggi badan dengan titik nol pada telapak kaki atau pada lantai saat berdiri. Jika sebuah meja panjangnya 12 kali panjang pensil, maka benda lain yang panjangnya 36 pensil adalah tiga kali panjang meja tersebut. Jadi penafsiran hasilnya bersifat rasio. Satuan ukurannya pun bisa dibalik. Misalnya panjang pensil dikatakan sebagai seperduabelas meja. Ada kalanya orang tak dapat menetapkan lokasi dari titik nol absolut untuk sebuah skala, karena atribut yang hendak diukur tidak kelihatan. Misalnya ketika hendak mengukur temperatur (suhu). Di sini, pengukuran dilakukan secara tidak langsung, yaitu dengan bantuan benda lain yang peka terhadap perubahan suhu, misalnya merkuri. Skala ukuran ditetapkan atas dasar pengamatan terhadap perubahan volume pada benda tersebut. Pada skala ini angka nol bukan berarti tak adanya kuantitas suhu.

Hampir semua pengukuran di bidang psikologi dilakukan tanpa skala dengan titik nol absolut. Seorang dengan skor nol pada sebuah tes Bahasa Inggris bukan berarti sama sekali tak memiliki pengetahuan Bahasa Inggris bahkan satu patah kata pun. Sebab itu, penafsiran hasil pengukuran tak dapat dilakukan secara rasio. Misalnya, seorang dengan skor 50 tidak dapat ditafsirkan sebagai dua kali lebih banyak pengetahuannya dibandingkan dengan orang yang mendapat skor 25. Skala ukuran yang tidak menggunakan titik nol absolut ini dinamakan skala interval. Dalam analisis statistika, tidak adanya titik nol absolut bukanlah merupakan halangan. Justru, kebanyakan metode statistika menggunakan angka rata-rata (*mean*) sebagai titik nol, bahkan pada data dengan skala rasio. Yaitu dengan cara mengubah variabel X menjadi $(X - \bar{X})$, sehingga langsung dapat diketahui apakah suatu nilai berada di bawah atau di atas angka rata-rata (*mean*) hanya dengan melihat tandanya, apakah positif atau negatif. Perubahan lokasi titik nol pada suatu skala dapat dilakukan dengan menambah atau mengurangi dengan suatu konstan pada setiap kasus. Hal ini tidak akan mengubah susunan distribusi frekuensi maupun indeks atau koefisien statistik seperti ranking, range, varians, korelasi, dan sebagainya. Yang terjadi hanyalah pergeseran lokasi dari keseluruhan distribusi frekuensi suatu variabel pada garis kontinum skala.

Kadang-kadang ada literatur yang menggunakan istilah skala ordinal atau bahkan sebutan skala nominal. Pada dasarnya penggunaan istilah ini tidak tepat karena dalam hal ini tidak terdapat *titik nol* dan *satuan ukuran*, yang menjadi syarat untuk sebuah skala. Oleh sebab itu tak dapat dinamakan skala ukuran. Apa yang disebut sebagai skala ordinal sebetulnya adalah upaya pengukuran atas suatu variabel kontinum dengan menggunakan kategori yang berjenjang (*ordered-categorical*). Misalnya kategori: baik sekali, baik, sedang, kurang, dan kurang sekali. Data yang dihasilkan adalah bilangan bulat positif (*integer*), yaitu frekuensi kejadian pada setiap kategori tersebut. Data seperti ini tak boleh dianalisis seakan ia bilangan riil pada skala interval. Namun demikian, kini telah tersedia metodologi untuk mentransformasikan data yang berupa frekuensi/proporsi pada setiap kategori yang berjenjang tersebut menjadi ukuran pada suatu skala interval. Justru inilah sebenarnya landasan logika dari sebagian besar pengukuran di bidang psikologi. Sebagian besar metode analisis data dengan statistika menggunakan asumsi bahwa data yang dianalisis

adalah hasil pengukuran dalam skala interval.

Dampak Kesalahan Pengukuran

Pada penelitian di bidang psikologi, sosiologi, pendidikan, ekonomi, kesehatan masyarakat, politik, dan ilmu sosial lainnya, salah satu ancaman terhadap kualitas analisis data jika menggunakan statistika ialah kesalahan pengukuran (*measurement error*). Koefisien statistik seperti korelasi, regresi, muatan faktor, dsb., akan mengalami "atenuasi" (lebih rendah dari yang semestinya/*underestimated*) karena alat dan prosedur yang digunakan dalam pengukuran tidak dapat menghasilkan data yang sepenuhnya valid dan reliabel. Ini terjadi karena pengukuran hanya dapat dilakukan secara tidak langsung karena atribut atau karakter yang hendak diukur umumnya berupa konsep atau konstruk yang tak dapat dilihat atau diobservasi secara fisik. Yang tampak dan terekam hanya indikatornya saja.

Kesalahan dalam pengukuran akan mengakibatkan diperolehnya data dengan ukuran variabilitas (varians) yang lebih besar dari pada yang semestinya. Misalkan ada 100 orang yang tingginya persis sama yaitu 160 cm (yaitu varians = nol), tetapi terukur dengan kesalahan. Andaikan kesalahan pengukuran yang terjadi bersifat random (tidak sistematis), maka sebagian orang akan terukur kurang dari 160 cm dan sebagian orang terukur lebih dari 160 cm, di mana kekurangan atau kelebihan dalam ukuran tersebut juga dapat bervariasi. Keadaan ini akan mengakibatkan diperolehnya data yang memiliki varians yang lebih besar dari nol, padahal seharusnya nol. Dengan kata lain, varians yang terlihat pada data sebenarnya adalah varians palsu, dan dinamakan varians kesalahan pengukuran. Selanjutnya, andaikan ada 100 orang yang tinggi badannya bervariasi dengan mean = 160 dan varians = 25 cm². Jika terjadi kesalahan pengukuran dan kesalahan itu bersifat random, maka data yang akan diperoleh akan memiliki mean yang sama yaitu 160 cm, tetapi akan memiliki varians yang jauh lebih besar dari 25 cm², misalnya 45 cm². Dalam hal ini, selisih antara varians pada data yang diperoleh (*observed variance*) dan varians yang sejatinya harus diperoleh jika tak terjadi kesalahan pengukuran (*true variance*), yaitu $45 \text{ cm}^2 - 25 \text{ cm}^2 = 20 \text{ cm}^2$, adalah merupakan variasi data karena kesalahan pengukuran, dan disebut varians kesalahan pengukuran (*Measurement Error Variance*). Varians kesalahan pengukuran ini berdampak negatif terhadap berbagai koefisien statistik ketika data hasil observasi dianalisis secara langsung tanpa koreksi. Seperti telah disebutkan sebelumnya, dapat terjadi atenuasi (*underestimated*) terhadap misalnya, koefisien regresi atau korelasi, sehingga suatu statistik yang sebenarnya signifikan menjadi dilaporkan sebagai tidak signifikan.

Setidaknya ada dua cara koreksi yang dapat dilakukan agar diperoleh koefisien yang terbebas dari atenuasi tersebut: (a) dilakukan estimasi terhadap *true scores* untuk setiap kasus (misalnya menggunakan *software* seperti BILOG, MULTILOG, Mplus, dsb.) kemudian *true scores* inilah yang dijadikan data dalam analisis; dan (b) menggunakan model statistika yang memperhitungkan dan mengkoreksi dampak dari *measurement error* terhadap setiap koefisien yang di-estimate (misalnya "confirmatory factor analysis")

dan "structural equation modeling" dengan software seperti LISREL, MPLUS, EQS, dsb).

Contoh 1: Koreksi dengan SEM

Structural Equation Modeling (SEM) merupakan metode statistika yang sangat generik karena dapat diterapkan pada semua jenis analisis statistika yang berbasiskan regresi. Di sini, LISREL adalah software yang paling populer di kalangan peneliti ilmu sosial. Berikut adalah ilustrasi dengan data riil (bukan simulasi) untuk menunjukkan betapa seriusnya dampak buruk dari *measurement error* dalam setiap analisis statistika.

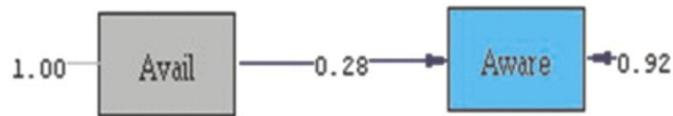
Misalkan seorang peneliti bidang kesehatan ingin mengetahui apakah kesadaran akan pentingnya makanan bergizi (*Awareness*) ditentukan oleh tingkat ketersediaan makanan di rumah (*Availability*). Ia berniat melakukan analisis regresi sederhana dengan variabel *Availability* ('Avail') sebagai *Independent Variable* dan *Awareness* ('Aware') sebagai *Dependent Variable*. Kedua variabel ini diukur menggunakan skala rating tipe Likert dengan lima kategori pilihan, mulai dari "sangat setuju" sampai dengan "sangat tidak setuju" pada setiap itemnya. Alat ukur masing-masing variabel terdiri dari lima item.

Secara tradisional, peneliti biasanya menjumlahkan skor setiap item sehingga diperoleh skor total untuk kedua variabel, lalu dilakukan analisis regresi. Hasilnya, diperoleh $R^2 = 0,08$ atau koefisien regresi (*standardized*) $\beta = 0,28$. Dalam bentuk diagram dapat dilihat pada Gambar-1. Selanjutnya data yang sama dianalisis dengan pendekatan SEM, menggunakan software LISREL. Di sini, analisis regresi dilakukan pada tingkat *true score* menggunakan model dengan struktur seperti pada Gambar 2 di bawah, yang selain memperhitungkan perbedaan reliabilitas dari setiap item, juga mengakomodasi saling korelasi antara "measurement error" pada item yang satu dengan lainnya. Hasilnya, ternyata diperoleh koefisien regresi (*standardized*) $\beta = 0,57$ dengan $R^2 = 0,33$. Ternyata dampak buruk dari *measurement error* cukup dramatis.

Pada sebuah penelitian yang penulis lakukan dengan sampel sebanyak 1855 orang, terdapat dua variabel yang dapat dijadikan sebagai ilustrasi dalam masalah ini, yaitu *emphaty* (EM diukur dengan 3 item), dan *task commitment* (TC diukur dengan 10 item). Semua itemnya adalah dalam bentuk skala Likert dengan 4 tingkatan pilihan. Pertama dilakukan analisis regresi dengan TC sebagai *dependent variable* dan EM sebagai *independent variable*, di mana keduanya diukur dengan skor total dari menjumlahkan skor item dan hasilnya dianggap sebagai ukuran berskala interval. Sebagai hasilnya diperoleh koefisien regresi (*standardized*, agar komparabel skalanya) sebesar 0,0989. Hal ini terlihat pada Gambar 3. Selanjutnya, dilakukan analisis regresi yang sama tetapi antar variabel laten EM dan TC di mana semua item masing-masing dijadikan indikator untuk mengukur (dengan analisis faktor). Sebagai hasilnya, diperoleh koefisien regresi (*standardized*) sebesar 0,33. Situasi ini dapat dilihat pada Gambar 4 jelas tampak di sini bahwa koefisien regresi sedemikian teratenuasi sehingga koefisien dalam ukuran *true score*

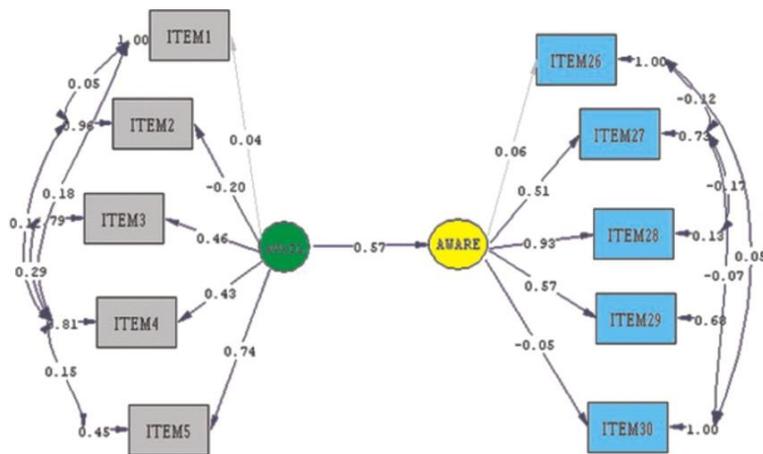
menghasilkan angka yang hampir tiga setengah kali lipat besarnya.

Ilustrasi di atas hanyalah pada analisis yang sangat sederhana. Bisa dibayangkan bagaimana seriusnya kesalahan yang mungkin terjadi pada model analisis yang lebih kompleks dengan banyak variabel (*multivariate*), mengingat dampak yang terjadi dapat saling berkait dan bersifat kumulatif. Terlebih lagi pada penelitian bidang psikologi, pendidikan, dan sosiologi, yang nyaris semua variabelnya adalah dalam bentuk konsep atau konstruk yang sangat abstrak (*latent variables*), sehingga validitas dan reliabilitas pengukurannya selalu merupakan masalah. Pada situasi ini, penggunaan metode seperti SEM dengan *software* seperti LISREL, Mplus, EQS, atau sejenisnya, kiranya merupakan suatu keharusan.



Gambar 1

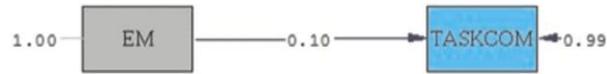
Koefisien Regresi menggunakan Data "Raw-Scores" Hasil Penjumlahan Skor Item



Gambar 2

Koefisien Regresi menggunakan True-Score

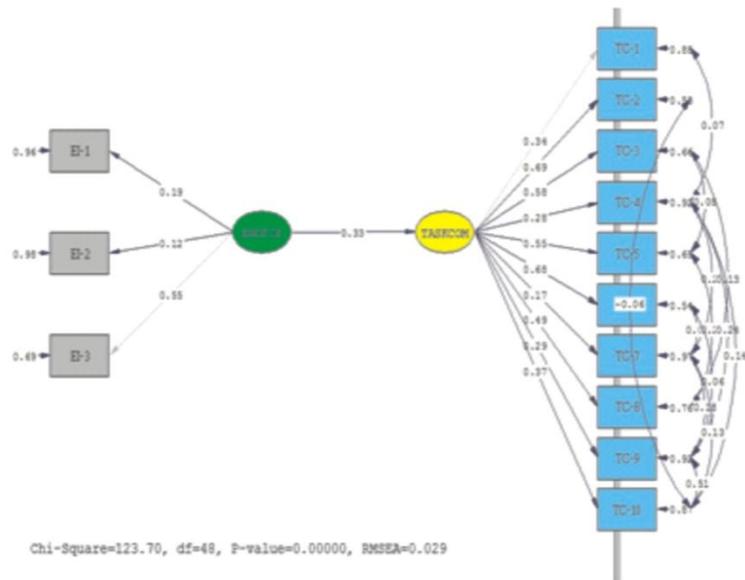
Contoh 2: Koreksi Dengan SEM



Chi-Square=0.00, df=0, P-value=1.00000, RMSEA=0.000

Gambar 3

Regresi dengan Raw Score



Chi-Square=123.70, df=48, P-value=0.00000, RMSEA=0.029

Gambar 4

Regresi dengan True Score

DAFTAR PUSTAKA

- Bentler, P. M. (2001). *EQS structural equation modelling software*. Multivariate Software, Inc. Los Angeles, CA.
- Joreskog, K.G., & Sorbom, D. (1996). *LISREL 8: User's reference guide*. Scientific Software International. Chicago, IL. www.ssicentral.com
- Muthen, L.K. & Muthen, B.O. (1998,2007). *Mplus user's guide*. Fifth edition. Los Angeles. CA: Muthen & Muthen. www.statmodel.com
- Thissen, D., Hung Chen, W., and Bock, R.D. (2003). *MULTILOG Scientific software international*. Chicago, IL www.ssicentral.com
- Zimowsky, M., Muraki, E., Mislevy, R., & Bock, R.D. (2003). *BILOG- MG. Scientific software international*. Chicago, IL. www.Ssicentral.com.