JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia), 9(2), 2020, 1-11
DOI: http://dx.doi.org/10.15408/jp3i.v9i2.XXXXX
http://journal.uinjkt.ac.id/index.php/jp3i

Uji Validitas Konstruk dengan CFA dan Pelaporannya

Jahja Umar, Yunita Faela Nisa
Fakultas Psikologi, UIN Syarif Hidayatullah Jakarta, Indonesia
yunita.faela@uinjkt.ac.id

Abstract

Confirmatory Factor Analysis (CFA) is the most reliable method of construct validity analysis in the fields of psychology, education and social sciences. From the author's observations on research articles as well as bachelor and graduate theses, and dissertations using CFA, it was found that there are a lot of misunderstandings and incompleteness in reporting CFA analysis. This paper is intended as an effort to improve this situation and provide recommendations in reporting data analysis using CFA. At the very least, this article is to show the important things that must be considered in understanding and using CFA and test theory in general.

***Keywords**: construct validity, confirmatory factor analysis, reporting a research.*

Abstrak

Confirmatory factor analysis (CFA) merupakan metode analisis yang sampai saat ini paling diandalkan dalam pengujian validitas konstruk suatu alat ukur di bidang psikologi, pendidikan, dan ilmu sosial pada umumnya. Dari pengamatan penulis terhadap artikel penelitian maupun skripsi, tesis, dan disertasi yang menggunakan CFA, masih ditemukan cukup banyak terjadi kekeliruan pemahaman maupun ketidaklengkapan dalam cara melaporkan hasil analisis CFA. Tulisan ini dimaksudkan sebagai upaya memperbaiki keadaan tersebut serta tentang bagaimana sebaiknya melaporkan hasil analisis data menggunakan CFA. Setidaknya, artikel ini untuk menunjukkan hal-hal penting yang harus diperhatikan dalam memahami dan menggunakan CFA dan teori tes pada umumnya.

Kata Kunci: uji validitas konstruk, *confirmatory factor analysis*, laporan penelitian.

Validitas Konstruk dan *Confirmatory Factor Analysis* (CFA)

Analisis Faktor adalah suatu metode analisis untuk menemukan apakah terdapat satu atau beberapa variabel yang bersifat *latent* (tak dapat diamati secara langsung) yang menjadi penyebab mengapa sehimpunan variabel saling berkorelasi. Istilah ini pertama kali dikemukakan oleh Spearman (1904) ketika ia berteori tentang adanya suatu variabel (*common factor*) yang menjadi penyebab mengapa sehimpunan skor hasil tes kemampuan kognitif saling berkorelasi, yang dinamakannya "*General Intelligence*". Selanjutnya, dengan metode ini, para ahli psikologi berhasil menemukan berbagai faktor (dimensi) dari variabel seperti "*basic abilities*", "*personality traits*", dan sebagainya. Nama-nama seperti Thurstone (1947), Cattell (1978), Guilford (1952), Comrey (1973) dan banyak lagi lainnya, merupakan orang-orang yang mengembangkan metode analisis faktor menjadi makin canggih. Namun demikian, selama sekitar 70 tahun metode analisis faktor tidak dianggap sebagai metode yang ilmiah oleh para ahli statistika karena setiap langkahnya yang bersifat subjektif. Seperti diketahui, analisis faktor pada waktu itu umumnya terdiri dari tiga langkah, yaitu (1) menentukan banyaknya faktor (*factor extraction*), (2) jika ditentukan lebih dari satu faktor, lalu dilakukan rotasi faktor (secara geometrik) untuk mendapatkan "*simple structure*" di mana suatu variabel hanya terkait dengan satu faktor saja, dan (3) menetapkan "nama" dari variabel *latent* (faktor) yang ditemukan. Pengambilan keputusan pada setiap langkah ini dilakukan tanpa adanya suatu kriteria yang objektif. Yang dilakukan adalah menghitung korelasi antarvariabel kemudian menghitung "*eigen value*" dari matriks korelasi tersebut, disertai dengan masing-masing "*eigen vector*"nya. Keputusan diambil berdasarkan "pertimbangan subjektif" terhadap besaran dan komposisi *eigen value* dan *eigen vector* tersebut. Untuk suatu data yang sama, setiap peneliti dapat menghasilkan keputusan yang jauh berbeda padahal prosedur dan metodenya sama. Juga tidak ada kegiatan "*statistical*" seperti uji hipotesis dengan tes signifikan, dsb.

Barulah di sekitar tahun 1970-an metode analisis faktor dipandang sebagai metode statistika ketika Lawley dan Maxwell (1971) dan Joreskog (1968,1969) mengemukakan model regresi, di mana: (1) variabel yang teramati (*observed variables*) dijadikan "*dependent variables*", yang nilainya bergantung kepada (dipengaruhi oleh) tinggi rendahnya nilai "*latent variables*" (*factors*) yang dijadikan "*independent variables*", dan (2) parameter-parameternya (koefisien regresi, korelasi antarfaktor, dan varians/kovarians residual) diestimasi dengan metode "*maximum likelihood*". Dalam hal ini, dapat dilakukan 2 (dua) kegiatan statistika, yaitu: (1) menguji hipotesis apakah model teori yang ditetapkan (di mana banyaknya faktor serta variabel yang digunakan untuk "mengukur" masing-masing faktor itu telah ditetapkan, adalah "*fit*" (sesuai) dengan data, dan (2) jika suatu model teoretis tentang faktor telah terbukti "*fit*" dengan data (dinyatakan diterima) maka dapat dilanjutkan dengan uji hipotesis (tes signifikan) terhadap setiap parameter dari model tersebut. Metode analisis faktor sebagai "model statistika" ini kemudian dikenal dengan sebutan "*Confirmatory Factor Analysis*" (CFA), sedangkan metode yang lama yang dianggap tidak ilmiah, disebut "*Exploratory Factor Analysis*" (EFA).

Karena sifatnya yang "konfirmatorik" maka CFA dapat digunakan untuk menguji validitas konstruk dari sebuah tes/alat ukur psikologi. Dengan CFA, bisa diuji (dikonfirmasi) sejauh mana seluruh *item* dari tes tersebut memang mengukur/memberikan informasi tentang satu hal saja, yaitu apa yang hendak diukur. Sebagai ilustrasi, misalkan ada sebuah tes "kemampuan verbal" yang terdiri dari 20 butir soal (*item*), yang berarti "diteorikan bahwa 20 butir soal tersebut semuanya hanya mengukur satu hal (faktor) saja" yaitu "kemampuan verbal". Artinya, jika memang teori ini benar maka seharusnya "model satu faktor" akan "*fit*" dengan data. Model satu faktor disebut juga "model unidimensional", dan setiap alat ukur harus memenuhi azas ini. Dalam sebuah alat ukur psikologi, semua *item*nya harus mengukur hanya satu hal saja yaitu konstruk yang hendak diukur. Jika ada satu atau sebagian *item*nya mengukur hal lain, maka berarti *item* tersebut tidak valid. Lalu, bagaimana cara

membuktikan secara empirik bahwa "model (teori) satu faktor" adalah sesuai (*fit*) dengan kenyataan (data)? Secara ringkas, logika dan prosedurnya adalah sebagai berikut:

1. Menetapkan spesifikasi dari model, yaitu mendeskripsikan (merumuskan) model (teori) secara verbal, lalu dirumuskan dengan (dibuat) gambar atau diagram, dan berdasarkan diagram itu dibuat rumus-rumus persamaan regresi, di mana setiap *item* diteorikan sebagai "*dependent variable*" sedangkan faktor (variabel *latent* yang hendak diukur) dijadikan "*independent variable*". Karena pada sebuah model pengukuran hanya ada satu faktor, maka pada contoh di atas akan ada 20 persamaan regresi sederhana, yang semuanya memiliki satu *independent variable* yang sama. Jika pada setiap persamaan regresi koesfisiennya disebut lambda (λ) dan varians dari "residual" disebut theta (θ), maka akan ada 40 parameter (yang nilainya tak diketahui) yaitu 20 buah λ dan 20 buah θ. Skala ukuran untuk variabel *latent* (faktor) ditetapkan dengan standardisasi (varians=1).
2. Karena yang menjadi tujuan adalah membuktikan apakah "*saling berkorelasinya"* 20 *item* hanya disebabkan oleh satu faktor saja, maka dibuatlah *"persamaan korelasi*" antara setiap pasangan *item*, dengan menggunakan 20 persamaan regresi tersebut di atas (termasuk persamaan korelasi antara suatu *item* dengan dirinya sendiri yang nilainya adalah satu). Dalam hal ini, setiap persamaan korelasi dinyatakan dalam simbol parameter yaitu λ dan θ. Karena matriks korelasi bersifat simetri, maka jumlah "persamaan korelasi" yang "harus dibuat rumusnya" adalah sebanyak p(p+1)/2 , di mana p adalah banyaknya *item*. Untuk contoh ini, akan ada sebanyak 20*21/2 = 210 buah persamaan korelasi, yang semuanya dinyatakan dalam bentuk simbol λ dan θ. Sebagai contoh, persamaan untuk korelasi antara *item* 2 dan *item* 1 adalah $\sigma_{21} = \lambda_2\lambda_1$, sedangkan persamaan korelasi *item* 4 dengan dirinya sendiri adalah $\sigma_{44} = \lambda_4^2 + \theta_{44}$, dst. (lebih detail dapat dilihat pada Umar, 2019). Jika seluruhnya dituliskan dalam bentuk matriks, 210 persamaan tersebut dapat ditulis dengan simbol $\Sigma = \Lambda\Lambda' + \Theta$ di mana Σ (sigma) adalah matrik korelasi antar*item* berdasarkan model (teori), Λ adalah matriks koefisien regresi yang disebut "*factor loadings*", dan Θ adalah matriks varians-kovarians antarresidual (namun di sini yang ada hanya varians θ saja, yang disebut varians dari "*measurement errors*"). Dengan kata lain, matriks Θ pada model unidimensional (contoh ini) adalah matriks yang bersifat diagonal.
3. Setelah model (perumusan teori secara matematis) ditetapkan, lalu dikumpulkan data (sampel) dari lapangan, untuk kemudian dihitung matriks korelasi antara 20 *item* yang ada, dan matriks korelasi berdasarkan data ini diberi simbol S. Jika apa yang diteorikan (model unidimensional) memang benar adanya, maka seharusnya **"tidak ada perbedaan"** antara Σ (matriks yang diramalkan oleh teori) dan S (matriks yang diperoleh dari data). Artinya, dibuat hipotesis nihil $S - \Sigma = 0$, atau dapat ditulis $S = \Sigma$. Persamaan $S = \Sigma$ ini terdiri dari 210 persamaan dengan 40 parameter yang tak diketahui nilainya (lihat butir-2). Dengan metode *"maximum likelihood"* dan menggunakan *software* seperti Lisrel (Joreskog dan Sorbom, 1993) atau Mplus (Muthen dan Muthen, 2017), nilai dari 40 parameter ini dapat diperoleh. Kegiatan ini disebut "estimasi parameter". Setiap *software* biasanya menyediakan juga berbagai metode estimasi lain, selain *maximum likelihood*.
4. Dengan menggunakan hasil estimasi (nilai yang diperoleh untuk 20 λ dan 20 θ), lalu dihitung matriks *"korelasi yang diharapkan oleh teori"*, yaitu Σ, dengan rumus pada butir-2 di atas, yaitu $\Sigma = \Lambda\Lambda' + \Theta$. Jadi matrik Σ itu tak ubahnya seperti halnya Y' (nilai Y hasil prediksi) pada analisis regresi biasa.
5. Menghitung selisih antara "matriks korelasi yang diharapkan oleh teori" dengan "matriks korelasi menurut data lapangan", untuk mengkonfirmasi apakah $S - \Sigma = 0$ (uji statistik apakah selisih

$S - \Sigma$ signifikan berbeda dari nol). Misalnya dengan uji Chi Square, atau dengan indeks lain seperti *Root Mean Square Error of Approximation* (RMSEA), *Comparative Fit Index* (CFI), *Tucker-Lewis Index* (TLI), dsb. Hasil hitungan dan uji statistik ini tersedia pada *output software* seperti Lisrel dan Mplus.

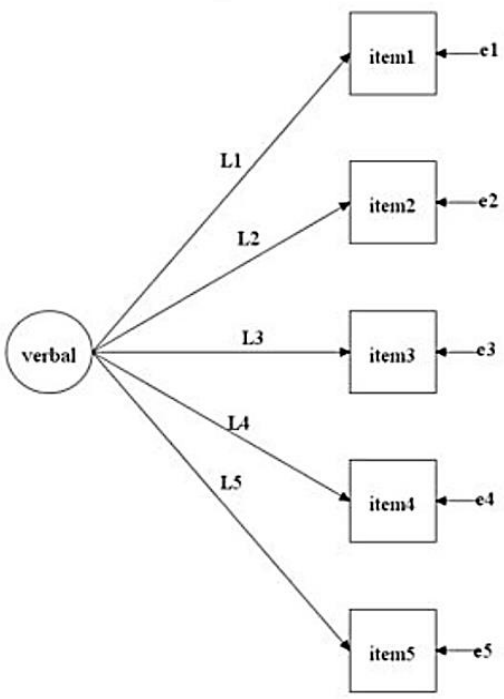
6. Jika selisih tersebut di atas secara statistik **TIDAK** signifikan, maka hipotesis nihil $S - \Sigma = 0$, **TIDAK** ditolak, yang berarti **model teori "satu faktor" *fit* dengan data** dan bisa diterima. Artinya, tes yang terdiri dari 20 *item* tersebut **VALID** mengukur hanya satu faktor, yaitu apa yang direncanakan untuk diukur.
7. Jika model diterima (valid) maka dapat dilanjutkan dengan uji signifikan terhadap setiap koefisien λ. Artinya, apakah masing-masing *item* memberikan kontribusi yang signifikan dalam mengukur faktor (konstruk) yang ditargetkan (di sini adalah kemampuan verbal). Hal ini bisa dilakukan dengan uji **t**-test, yang juga tersedia pada *software* seperti Lisrel atau Mplus.
8. Jika model teori (unidimensional) yang diuji **TIDAK fit** dengan data alias **ditolak**, maka dapat dilakukan modifikasi model misalnya dengan men*drop item* tertentu yang menjadi penyebab tidak *fit*nya model satu faktor. Untuk menemukan *item* mana yang perlu didrop, dapat dilakukan dengan menambah parameter, misalnya dengan membebaskan korelasi antara kesalahan pengukuran (elemen nondiagonal dari matriks Θ), sampai *model fit* tercapai, dan kemudian *item* dengan residual yang banyak korelasinya dapat di*drop* sehingga akhirnya dapat diperoleh model unidimensional yang *fit* dengan data.

Langkah-langkah untuk uji validitas konstruk seperti disajikan di atas, dapat dilakukan untuk semua model CFA, termasuk model multimensional di mana terdapat banyak faktor, baik yang saling berkorelasi satu sama lain maupun yang tidak (*orthogonal*). Untuk yang saling berkorelasi, dapat dilanjutkan dengan model "*second order* CFA". Juga bisa untuk menguji model yang disertai berbagai "constraints" pada parameternya. Misalnya apakah semua *item* dalam sebuah tes bersifat "*parallel*" (seperti pada teori tes klasik dan Rasch Model), dan sebagainya. Dalam bentuk yang lebih canggih, CFA juga dapat digunakan untuk mengatasi situasi di mana beberapa *item* mengandung "bias". Artinya, ada beberapa *item* yang sebenarnya valid dan signifikan dalam mengukur apa yang hendak diukur namun pada saat yang sama mengukur juga hal lain. Misalnya butir soal matematika dalam bentuk soal cerita, di mana selain mengukur matematika juga mungkin mengukur kemampuan memahani bacaan. Dalam hal ini, orang bisa menjawab salah karena ia kurang memahami apa yang ditanyakan. Pada pengukuran psikologi keadaan ini sangat sering terjadi, misalnya *item* untuk mengukur "kerja sama" ternyata justru mengukur "*altruism*", atau *item* yang diniatkan untuk mengukur "sikap" tapi juga mengukur "pengetahuan", dsb. Adanya *item* yang mengandung bias akan menyebabkan model unidimensional tidak *fit* dengan data. Di sini, pemodelan dengan CFA tidak saja dapat digunakan untuk mendeteksi *item* bias, tetapi juga dapat digunakan untuk mengkoreksinya tanpa harus men*drop item* yang bersangkutan. Metode CFA saat ini telah menjadi suatu metode yang sangat "*central*" peranannya di dalam psikometrika. Lebih dari itu, CFA juga merupakan suatu metode yang bersifat "*generic*" di mana banyak model statistika lain yang sebenarnya merupakan salah satu bentuk saja dari CFA. Boleh dikatakan, semua teknik analisis yang berkaitan dengan hubungan antara variabel *latent* dan variabel "*observed*", adalah suatu bentuk CFA. Misalnya, CFA dengan "c*ategorical observed variable*" dinamakan "*Item Response Theory*". Atau jika "*latent variable*"-nya yang bersifat kategorikal maka CFA dinamakan "*Latent Class Analysis*" (LCA), dan CFA yang korelasi antar*variable latent* "dimodel" menjadi analisis regresi lalu diberi nama "*Structural Equation Modeling*" (SEM), dan sebagainya.

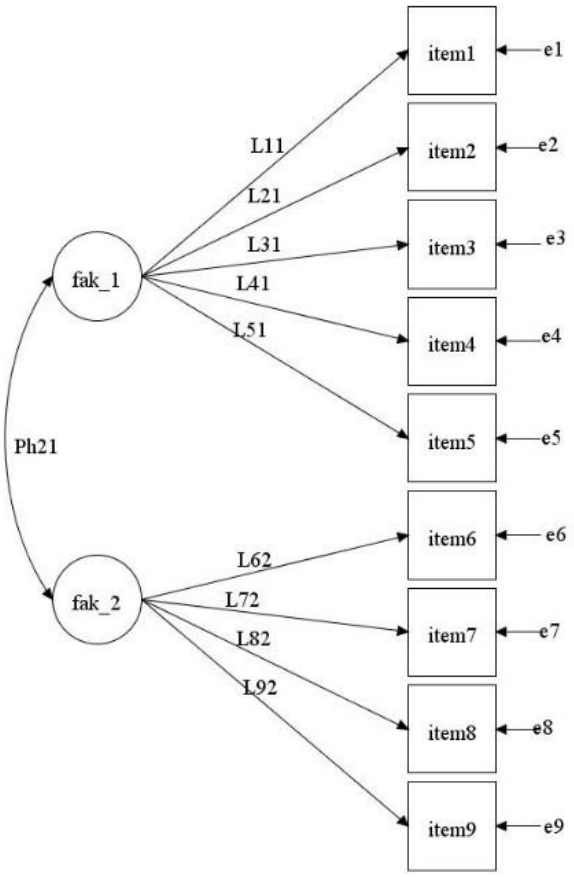
Pelaporan Hasil CFA

Apa yang perlu diperhatikan dalam melaporkan hasil analisis CFA? Dari pengamatan penulis terhadap artikel yang terbit di jurnal ini maupun beberapa jurnal internasional, terlihat adanya variasi yang cukup besar tentang apa yang dilaporkan. Bahkan kadang terlihat bahwa konsep CFA itu sendiri belum dipahami dengan baik oleh penulisnya. Oleh sebab itu tulisan ini dimaksudkan untuk menambah rujukan dan bahan pertimbangan dalam memutuskan apa yang perlu dimasukkan dalam pelaporan hasil CFA. Tulisan ini bukan yang pertama karena sebelumnya juga sudah terbit artikel seperti ini, meskipun mungkin berbeda penekanannya. Hal ini dapat dilihat pada artikel oleh misalnya, Jackson, et.al. (2009), Schreiber, et.al. (2006), atau McDonald and Moon-Ho (2002).

Sebetulnya, apa yang perlu dilaporkan dalam hasil analisis dengan CFA adalah terkait dengan langkah (prosedur) CFA seperti telah diuraikan di bagian sebelumnya. Misalnya pada butir ke satu, yang dikenal dengan istilah "*model specification*". Pada setiap penggunaan CFA perlu dilaporkan deskripsi dari model secara verbal (uraian), yang biasanya diikuti oleh deskripsi secara figural (diagram), dan kadang-kadang diperlukan juga deskripsi model secara matematikal (persamaan-persamaan). Dalam hal ini, uraian verbal diharuskan ada, lalu diagram sebaiknya ada agar pembaca lebih mudah memahami apa yang dikonsepkan (diteorikan), sedangkan rumusan matematisnya diperlukan tergantung siapa yang diharapkan menjadi pembacanya. Kebanyakan laporan hasil CFA menyertakan sebuah gambar (diagram) di mana faktor (konstruk) yang hendak diukur digambarkan dengan lingkaran, sedangkan "*observed*" variabel atau *item* dengan gambar kotak, lalu dibuat tanda panah dari faktor menuju *item*, dan panah yang pendek dari variabel residual (kesalahan pengukuran) yang juga menuju *item*. Cara membuat diagram seperti ini telah merupakan kebiasaan umum dalam CFA, yang pertamanya dipopulerkan dalam Lisrel Versi 4 (Joreskog dan Sorbom, 1981). Gambar-1 berikut adalah contoh diagram untuk model pengukuran kemampuan Verbal dengan 5 (lima) *item*, di mana L1 sampai L5 adalah koefisien muatan faktor (λ) dan e_1 sampai e_5 adalah variabel residual (termasuk di sini kesalahan pengukuran) yang variannya adalah disebut θ. Gambar 2 adalah contoh diagram untuk model dua faktor.



Gambar 1. Model Pengukuran Kemampuan Verbal dengan Lima *Item*



Gambar 2. Diagram Model Dua Faktor

Untuk model CFA multidimensional perlu digambar pula panah melemgkung dengan dua mata panah di ujungnya untuk menunjukkan korelasi antarfaktor. Hal yang sama juga digunakan untuk menunjukkan korelasi antardua residual pada 2 (dua) *item*. Adapun persamaan-persamaan matematik yang menggambarkan model yang sedang diuji secara empirik, sebaiknya juga ditampilkan jika diterbitkan sebagai artikel di bidang metodologi psikometrika (bukan aplikasinya). Pada pelaporan hasil penggunaan CFA untuk tujuan uji validitas instrumen, biasanya tidak memerlukan ditampilkannya rumus matematika yang lengkap, apalagi yang terkait dengan pembuktian dan perhitungannya.

Karena dalam menguji kebenaran dari suatu model CFA (misalnya unidimensional) diperlukan estimasi untuk setiap parameter, yang dalam hal ini menggunakan data empiris, maka perlu dilaporkan populasi, sampel, dan teknik pengambilan sampel yang dilakukan. Hal ini penting bukan saja untuk kepentingan generalisasi dari hasil analisis tetapi yang tak kalah pentingnya adalah sejauh mana hasil estimasi parameter bisa dipercaya. Metode estimasi yang berbeda bisa membutuhkan besaran sampel (*sample size*) minimum yang berbeda pula. Dalam inferensi statistika, besaran sampel ini penting karena terkait dengan "*statistical power*" dari suatu hasil analisis. Selain terkait dengan metode estimasi parameter yang digunakan, besaran sampel yang dibutuhkan juga dipengaruhi oleh banyaknya parameter yang akan diestimasi. Selanjutnya, bentuk pemodelan statistika yang digunakan juga ikut mempengaruhi. Misalnya, dalam analisis regresi berganda ada semacam "*rule of thumb*" bahwa besaran sampel sebaiknya jangan kurang dari 20 kali banyaknya variabel prediktor (Pedhazur, 1997). Untuk penelitian eksperimental dan survei, bahkan diusulkan berbagai rumusan matematis tentang cara menghitung sampel minimum tersebut untuk tingkatan "*statistical power*" dan "*margin of error*" yang ditetapkan. Bagaimana dengan model yang kompleks seperti CFA? Materi ini tidak menjadi pokok bahasan dalam tulisan ini, namun untuk mudahnya penulis merujuk Joreskog saja kepada hasil studi simulasi oleh Bentler (1995) misalnya, di mana disarankan agar sampel yang digunakan sekurangnya 5 (lima) kali dari jumlah parameter yang

ada pada sebuah model CFA jika metode estimasinya adalah "*maximum likelihood*", atau 10 kali jika dengan estimator lain. Terkait pelaporan hasil CFA, yang penting harus dilaporkan adalah (1) besaran sampel dan cara pengambilan sampel, (2) metode estimasi yang digunakan untuk mendapatkan nilai parameter, dan (3) *software* yang digunakan untuk perhitungannya. Tentu saja sebaiknya disertai dengan argumen yang tepat. Semua ini terkait dengan langkah butir 2 dan butir 3 pada bagian sebelumnya.

Selajutnya, yang paling penting dan "*crucial*" dalam CFA adalah indeks *model fit* yang digunakan untuk memutuskan apakah model yang diuji diterima atau ditolak. Masih banyak permasalahan yang belum tuntas dalam hal ini dan sifatnya pun kontekstual. Pada dasarnya, dari semua indeks yang ada saat ini, hanya satu yang bersifat "*statistical test*" dan paling sesuai dan tepat untuk tujuan yang bersifat "konfirmatorik" yaitu uji "chi-square" (χ^2). Tak ada perdebatan tentang hal ini. Namun indeks ini banyak "digugat" karena cenderung membuat suatu model selalu ditolak sebab ia sangat sensitif terhadap besaran sampel. Penjelasannya adalah sebagai berikut. Nilai χ^2 diperoleh dari perkalian antara nilai fungsi minimum dengan besarnya sampel yang digunakan (N). Artinya, makin besar N akan makin besar pula nilai χ^2 yang berarti makin besar kemungkinannya "signifikan" untuk nilai *df* tertentu (model dinyatakan tidak *fit* dan ditolak). Padahal, yang dijadikan data untuk estimasi parameter CFA adalah matriks korelasi (bukan data individual), sehingga derajat kebebasan (*df*) dari χ^2 jauh lebih kecil. Perhitungan *df* dalam hal ini adalah banyaknya korelasi (data) dikurangi banyaknya parameter. Misalkan sebuah model CFA unidimensional dengan 10 *item*, banyaknya korelasi adalah (10*11)/2 = 55, sedangkan banyaknya parameter ada 20 buah (10 λ dan 10 θ), jadi derajat kebebasan (*df*) model ini adalah 55 – 20 = 35. Kalau misalnya besaran sampel 1.000 orang maka model ini hampir pasti akan ditolak karena χ^2 yang signifikan, namun kalau besarnya sampel hanya 100 orang maka kemungkinan besar model yang sama akan diterima (χ^2 nonsignifikan). Di sisi lain, hasil estimasi (apalagi *maximum likelihood*) akan lebih bisa dipercaya jika sampelnya besar. Jadi ada semacam dilema. Oleh sebab itu para ahli statistika kemudian mengembangkan berbagai macam indeks *model fit* yang relatif tak dipengaruhi besaran sampel. Sampai saat ini sedikitnya ada 25 jenis indeks *model fit*, namun hampir semuanya bukanlah uji statistik alias tidak bersifat probabilitas dalam penentuan signifikan tidaknya perbedaan antara S dan Σ. Akibatnya keputusan tentang model *fit* tidak lagi bersifat konfirmatorik. Dari semua alternatif indeks yang ada, terdapat satu yang masih bersifat uji statistik yaitu RMSEA dengan "*confident interval*" serta probabilitasnya. Indeks ini memang merupakan fungsi langsung dari χ^2 namun dengan koreksi terhadap dampak dari N pada suatu model dengan ***df*** tertentu. Oleh sebab itu penulis merekomendasikan RMSEA dalam uji validitas konstruk, yang kegiatannya bersifat konfirmatorik. Sedangkan indeks lain terutama CFI dan TLI, menurut pengalaman penulis bisa digunakan sebagai "*cross check*" jika ada hal yang "mencurigakan" pada hasil dan proses estimasi. Misalnya ketika RMSEA menunjukkan *model* fit namun CFI dan TLI nilainya kurang dari 0,9 (jika *model fit* maka indeks ini mestinya mendekati satu), bisa mengindikasikan adanya masalah dalam perhitungan estimasi (misalnya tidak konvergen, dsb.). Namun jika tujuan CFA adalah untuk melihat eksistensi dari suatu kostruk (sifatnya pengembangan teori) dan bukan untuk konfirmasi validitas suatu instrumen, memang bisa dimengerti jika tidak harus sepenuhnya "konfirmatorik" karena bersifat "*model generation*" (Joreskog et. al., 2016). Dalam hal ini penentuan "*model fit*" dengan indeks yang tidak berbasis populasi (*nonstatistical*) bisa dimaklumi.

Dari pembahasan di atas, pelaporan uji *model* fit untuk CFA terutama dalam rangka uji validitas konstruk, haruslah disertai nilai χ^2 dengan *df* dan probabilitas (signifikansi)nya. Jika model tidak *fit* maka dilaporkan nilai RMSEA sebagai alternatifnya. Namun perlu dicatat bahwa RMSEA pun tidaklah sama sekali terbebas dari pengaruh besaran sampel. Jika jumlah *item*nya sedikit tetapi sampel sangat besar (ribuan) maka RMSEA pun akan mengakibatkan model ditolak seperti halnya pada χ^2. Dalam hal ini,

pendekatan *Item Response Theory* di mana pola response (bukan N) yang dijadikan landasan perhitungan *df* dalam rangka uji signifikansi χ^2, dapat dilakukan karena respons terhadap *item* selalu menghasilkan data kategorikal. Sayangnya, alternatif ini hanya tersedia pada software tertentu. Misalnya pada software Mplus, hal ini dapat dilakukan dengan menggunakan estimator MLR (Muthen and Muthen, 2017). Namun jika jumlah *item*nya terlalu banyak, estimator inipun tak akan menghasilkan nilai χ^2 yang terpercaya. Ketika melaporkan uji *model fit* dengan RMSEA, agar disajikan nilai RMSEA dan "*confidence interval*" serta probabilitasnya. Dalam hal ini, dengan menggunakan taraf signifikansi 5%, model dinyakan fit dengan data jika (1) nilai RMSEA kurang dari 0,05, dan (2) dengan "*confidence interval*" yang termsuk di dalamnya nilai 0,05, serta (3) "probabilitas bahwa nilai RMSEA < 0,05*",* adalah lebih besar dari 5% (p > 0,05). Menurut penulis, paling tidak jika dua dari ketiga hal ini terpenuhi maka model dapat dinyatakan *fit*. Adalah lebih baik jika disebutkan juga nilai CFI dan TLI. Tak perlu melaporkan semua indeks lain yang sebetulnya bukan uji statistik walaupun mungkin bisa bermanfaat untuk tujuan yang nonkonfirmatorik, misalnya untuk studi yang oleh Joreskog (Joreskog et. al., 2016) disebut bersifat "*model generation*".

Walaupun estimasi parameter telah dilakukan sebelum uji *model fit*, namun pelaporan uji signifikansi atas setiap parameter haruslah "hanya" sesudah diperoleh *model fit*. Karena jika suatu model teori ditolak (tidak *fit*) maka tentu tak relevan untuk menguji hipotesis tentang parameternya. Setelah terkonfirmasi bahwa model unidimensional diterima (*fit* dengan data), langkah selanjutnya adalah melaporkan nilai parameter yang diperoleh serta uji signifikansinya. Biasanya dilaporkan berupa tabel koefisien muatan faktor untuk setiap *item*, yang disertai *standard error*, nilai z atau t, dan probabilitas (signifikansi)nya. Seperti lazimnya, nilai ***t*** atau ***z*** (absolut) yang lebih besar dari 1,96 berarti signifikan pada taraf 5%. Sebaiknya yang dilaporkan di sini adalah koefisien yang "*standardized*" agar besarannya bisa diperbandingkan secara langsung. Namun perlu dicatat bahwa tidak semua *software* CFA yang ada memiliki kapasitas dalam menghitung *standard error* (dan uji signifikansi) untuk koefisien muatan faktor yang "*standadized*". Dalam hal ini perlu ditampilkan nilai koefisien muatan faktor baik yang "unstandardized" (dengan uji signifikansinya) maupun yang "*standardized*". Tabel 1 adalah contoh untuk tes dengan lima *item* dan Tabel 2 untuk contoh model CFA dua faktor dengan sembilan *item*.

Tabel 1. Koefisien Muatan Faktor

Variabel	Estimates	S.E.	t	prop
Item-1	0,539	0,024	22,458	0,000
Item-2	0,677	0,019	34,831	0,000
Item-3	0,621	0,021	30,119	0,000
Item-4	0,591	0,021	28,109	0,000
Item-5	0,664	0,019	34,360	0,000

Tabel 2. Koefisien Muatan Faktor

Variabel	Faktor-1	Faktor-2	S.E.	t	prop
Item-1	0,539		0,024	22,458	0,000
Item-2	0,677		0,019	34,831	0,000
Item-3	0,621		0,021	30,119	0,000
Item-4	0,591		0,021	28,109	0,000
Item-5	0,664		0,019	34,360	0,000
Item-6		0,807	0,008	100,324	0,000
Item-7		0,543	0,012	45,405	0,000
Item-8		0,856	0,007	114,821	0,000
Item-9		0,764	0,009	89,289	0,000

Keterangan: Korelasi faktor-1 dan faktor-2 = 0,398, p < 0,05

Selain dalam bentuk tabel, ada juga yang melaporkan koefisien muatan faktor dengan cara menampilkan diagram dengan nilai koefisien pada setiap garis panahnya, dan jika tidak signifikan maka panahnya dihapus. Menurut pendapat penulis, penyajian dengan tabel akan lebih mudah dibaca dan dipahami. Adapun tabel yang berisi varians residual agak jarang dilaporkan kecuali pada konteks tertentu di mana varians tersebut akan dibahas. Jika model CFA yang diuji bersifat multifaktor, maka tabel korelasi antarfaktor biasanya juga ditampilkan. Dalam melaporkan koefisien muatan faktor seperti pada Tabel 1 atau Tabel 2, kadang-kadang ada penulis yang mengganti label nomor *item* dengan isi kalimat pertanyaan atau pernyataan pada setiap *item*. Jika memungkinkan tentu hal ini lebih baik.

Dalam uji validitas konstruk dengan CFA sering kali ada *item* yang harus di*drop* agar diperoleh model pengukuran (unidimensional) yang *fit* dengan data. Adapun kriteria untuk menentukan apakah sebuah *item* harus di*drop* adalah: (1) jika *item* memiliki koefisien muatan faktor yang negatif, ***atau*** (2) jika residual suatu *item* berkorelasi dengan banyak residual pada *item* lainnya, ***atau*** (3) jika koefisien muatan faktor tidak signifikan. Ketika terdapat beberapa *item* yang harus di*drop* (berarti *item* tidak valid), akan sangat baik jika disertai interpretasi tentang kemungkinan apa yang menyebabkan *item* tersebut tidak lulus uji validitas. Hal ini dapat dijadikan umpan balik bagi proses perbaikan atau penulisan butir soal. Suatu perangkat tes atau instrumen yang telah terbukti bersifat unidimensional (model satu faktor *fit* dengan data) dapat dinyatakan sebagai tes yang valid untuk mengukur konstruk yang ditargetkan. Artinya, skor yang dihasilkan secara logis akan menggambarkan tinggi rendahnya nilai pada konstruk atau atribut yang diukur. Karena tidak ada *item* yang mengukur atribut atau konstruk lain di dalam tes tersebut. Namun masalah yang kemudian timbul adalah prosedur *skoring*nya.

Meskipun semua *item* dalam suatu tes telah terbukti mengukur atribut yang sama (valid), tetapi selalu terdapat variasi dalam hal karakteristik *item* seperti tingkat kesukarannya, kemampuannya dalam membedakan individu satu dengan lainnya (*discriminating power*), dan tingkat variasi kesalahan pengukuran (*varians residual*) yang dimilikinya. Sebagai ilustrasi sederhana misalkan sebuah tes pengetahuan matematika yang terdiri dari sehimpunan *item* yang sudah pasti memiliki tingkat kesukaran yang berbeda-beda. Jika perbedaan ini tidak diperhitungkan dalam *skoring* (semua *item* dianggap sama atau paralel padahal sebenarnya tidak), maka skor yang dihasilkan akan menyesatkan dan tidak valid. Karena jika ada dua orang mendapatkan skor yang sama maka tidak bisa ditafsirkan bahwa kemampuan mereka sama. Ada dua pilihan yang tersedia dalam mengatasi masalah ini. Pertama adalah cara penskoran yang memperhitungkan perbedaan (variasi) antar*item*, yang artinya setiap *item* diberi bobot berdasarkan karakteristiknya. Jika cara ini yang ditempuh maka timbul pertanyaan bagaimana cara pembobotannya? Yang kedua adalah mengabaikan saja perbedaan antar*item* dan dianggap saja semua *item* dalam tes itu paralel. Pada cara inipun timbul pertanyaan: bagaimana cara membuktikan bahwa memang semua *item* memiliki karakteristik yang sama (paralel)? Artinya, perbedaan yang terlihat hanyalah karena fluktuasi random sampling dan tidak signifikan sehingga boleh diabaikan. Ternyata metode CFA dapat digunakan sebagai solusi kedua pertanyaan ini. Untuk yang pertama, agar setiap *item* dipertimbangkan karakteristiknya maka penskoran dilakukan berdasarkan pola jawaban (*response patterns*) pada butir soal. Misalkan ada tiga butir soal pilihan ganda yang diskor satu jika jawaban benar dan nol jika jawaban salah. Untuk tiga *item* akan ada $2^3 = 8$ kemungkinan pola jawaban yaitu (0 0 0), (1 0 0), (0 1 0), (0 0 1), (1 1 0), (1 0 1), (0 1 1), dan (1 1 1), yang setiap pola menghasilkan satu skor yang berbeda. Bayangkan jika tes terdiri dari 40 *item* maka banyaknya pola jawaban atau skor yang bisa dihasilkan oleh tes itu ada sebanyak $2^{40} = 1.099.511.627.776$ pola. Tentu saja hasilnya sangat cermat karena kemungkinan ada dua orang mendapat skor yang sama amatlah kecil yaitu kurang dari sepersatu triliun. Artinya, bahkan jika seluruh penduduk dunia menempuh tes itu pun nyaris tak akan ada yang skornya sama (kecuali saling contek). Jika terjadi, hampir pasti karena pembulatan, dan mengatasinya cukup ditambah saja desimalnya dan akan didapat skor yang berbeda. Cara skoring seperti ini dapat dilakukan dengan

metode "CFA *for Categorical Variables*" di mana perbedaan *item* dalam hal tingkat kesukaran (*thresholds*), daya pembeda (*factor loadings*), dan varians kesalahan pengukuran (*residual variances*) dapat diperhitungkan. Skor yang dihasilkan dikenal dengan sebutan "*true scores*" atau "*factor scores*". Namun tidak semua perangkat lunak CFA dapat menghasilkan skor ini. Selain itu juga terdapat berbagai metode untuk mendapatkannya. Salah satu yang paling canggih dalam perhitungan (estimasi) skor seperti ini ialah *software* Mplus (Muthen and Muthen, 2017). Cara skoring dengan CFA tak perlu dilaporkan jika yang dilakukan hanya terbatas pada uji validitas instrumen.

Yang memerlukan pelaporan khusus dengan CFA ialah jika skoring dilakukan dengan alternatif kedua, yaitu semua item dianggap paralel dan skor setiap orang dihitung dalam bentuk penjumlahan jawaban pada setiap *item*. Dalam hal ini, paralelitas *item* dijadikan asumsi yang harus diuji kebenarannya dengan CFA setelah asumsi unidimensionalitas terpenuhi. Artinya, perlu dilakukan pengujian model CFA yang unidimensional dan paralel. Apanya yang paralel? Idealnya adalah paralel dalam (1) tingkat kesukaran, (2) daya pembeda soal, dan (3) varians kesalahan pengukuran (residual). Dalam "CFA *for Categorical Variables*", ini berarti menguji *model fit* untuk sebuah model teoretis yang (1) semua nilai *Threshold* (tau, τ)nya sama, (2) semua nilai muatan faktor (λ) sama, dan (3) semua *varians residual* (θ) sama. Jika model seperti ini *fit* dengan data, berarti perbedaan yang ada pada tiga parameter tersebut hanyalah karena fluktuasi sampel dan tidak signifikan, sehingga boleh diabaikan dalam perhitungan *skoring*. Namun masalahnya model yang penuh dengan "restriksi" seperti itu jarang bisa dijumpai dalam kenyataan (data). Artinya, kalaupun diperoleh *model fit* biasanya setelah sangat banyak *item* yang harus di*drop* terlebih dahulu. Oleh sebab itu para ahli psikometrika (misalnya Lord dan Novick, 1968) menganggap bahwa jika satu saja yang paralel yaitu "*factor loading*", sudah cukup untuk menyatakan bahwa sebuah tes terdiri dari butir-butir *item* yang paralel. Alasannya, jika koefisien muatan faktor paralel maka berarti setiap *item* akan menghasilkan "*predicted*" atau "*expected true score*" yang sama, sehingga sudah cukup kuat untuk justifikasi penggunaan "*raw score*" (hasil penjumlahan skor *item* tanpa pembobotan). Tes yang memenuhi syarat unidimensional dan "*parallel factor loading*" ini dinamakan tes yang bersifat "*tau-equivalence*", boleh menggunakan "*raw score*" dan dilaporkan reliabilitasnya (misalnya Cronbach α). Dalam kaitannya dengan pelaporan hasil CFA, berarti untuk model yang lebih "*restricted*" ini perlu disampaikan (1) hasil uji *model fit* dengan nilai χ^2, *df* dan probabilitasnya, serta nilai RMSEA dengan "*confidence interval*" dan probabilitasnya, dan (2) besaran nilai koefisien muatan faktor yang berlaku untuk semua *item* dengan hasil uji signifikansinya. Seperti pada uji unidimensionalitas, interpretasi mengapa sebagian *item* harus di*drop* dalam proses ini pun sebaiknya dilaporkan. Tes yang dihasilkan biasanya lebih singkat dan cara penskorannya praktis sehingga bisa dilakukan oleh siapapun pengguna tes, tanpa diperlukannya pengetahuan psikometrika ataupun perhitungan *true* skor yang rumit. Seharusnya semua tes psikologi yang di dalam praktiknya menggunakan "*raw score*", harus melalui uji unidimensionalitas dan paralelitas seperti ini, dan melaporkan hasil uji validitas konstruk maupun uji paralelitas *item* tersebut dalam buku manual atau "*technical report*" yang menyertainya.

Daftar Pustaka

Bentler, P. M. (1995). *EQS – Structural equations program manual.* Multivariate Softwares. Encino. CA.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences.* Plenum Press. New york.

Comrey, A. L. (1973). *A first course in factor analysis*. Academic Press. New York.

Guilford, J.P. (1952). When not to factor analyze. *Psychological Bulletin,* 49, 26–37.

Jackson, D.L., Gillaspy, Jr., J. A., and Stephenson, R. P. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological Methods*, Vol. 14, No. 1, 6–23

Joreskog, K. G. (1968). New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology,* 21, 86–96.

Joreskog, K. G., (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* 34, 183–202.

Joreskog, K. G. And Sorbom, D. (1981). *LISREL IV. Linear structural relatioship by the method of maximum likehood. a user guide.* Scientific Softwares International. Chicago.

Joreskog, K. G. And Sorbom, D. (1993). *LISREL 8: Structural equations modeling with the simplis command language.* Scientific Softwares International. Chicago.

Joreskog, K. G., Olsson, U. H. And Wallentin, F. Y. (2016). *Multivariate analysis with LISREL*. Springer. Switzerland.

Lawley, D.N. and Maxwell, A. E. (1971). *Factor analysis as a statistical method. 2nd ed.* Butterworth. London.

Lord, F.M. and Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley. New York.

McDonald, R.P., and Ho, M. R. (2002), Principles and practice in reporting structural equation analyses. *Psychological Methods*, Vol. 7, No. 1, 64–82.

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus user's guide. Eighth edition.* Los Angeles, CA: Muthén & Muthén.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research. 3rd Ed*. Wadsworth Thomson Learning. New York.

Schreiber, J. B., Nora, A., Stage, F.K., Barlow, E.A., King, J. (2006) Reporting structural equation modeling and confirmatory factor analysis results: a review, *The Journal of Educational Research*, Jul. - Aug., 2006, Vol. 99, No. 6 (Jul-Aug., 2006), pp. 323–337.

Spearman, C. (1904). General intelligence: Objectively determined and measured. *American Journal of Psychology*, 5, 201–293.

Thurstone, L.L. (1947). *Multiple factor analysis.* University of Chicago Press. Chicago.