ANALISIS FAKTOR 2-LEVEL DALAM MODEL PERSAMAAN STRUKTURAL

Mohamad Waluyo, Abdurakhman, Zulaela
Pendidikan Matematika UMS, Matematika FMIPA UGM
Mohammad.waluyo@ums.ac.id, rachmanstat@ugm.ac.id, zulaela@ugm.ac.id

Abstrak

Model persamaan struktural (SEM) sangat berguna untuk mengetahui apakah suatu model antar variabel laten, variabel indikator fit dengan data yang diperoleh. Ketika data yang diperoleh cukup besar dan berhirarki maka diperlukan analasis penyesuaian dalam model struktural tersebut. Penyesuaian tersebut adalah dengan memasukkan variabel random efek dari tiap tingkatan hirarki. Variabel random untuk 2 level hirarki yaitu variabel random efek level-1 dan level-2. Metode estimasi yang digunakan dalam penelitian ini adalah maksimum likelihood (ML). Penggunaan metode ML ini merupakan metode yang umum dipakai dalam estimasi parameter dalam SEM namun mengalami kendala ketika terdapat data hilang. Sehingga algoritma ekspektasi-maksimasi (EM) diperlukan untuk mengestimasi data hilang.

Kata kunci: *SEM; multilevel; ML; algoritma EM.*

1. PENDAHULUAN

Model persamaan struktural telah lama dikenal dan berkembang sebagai salah satuanalisis multivariat. Bollen(1989) dalam Ghozali (2008) yang menyatakan *Structural Equation Modelling* (SEM) dapat menguji bersama-sama: (1) Model struktural yaituhubungan konstruk yang independen dan dependen, (2) Model pengukuran yaitu hubungan antara indikator dengan variabel laten.

Data berkelompok yaitu kumpulan data dengan peubah respon diukur sekali untuksetiap subjek (unit analisis) dan unit analisis dikelompokkan ke dalam, atau tersarangdi dalam atau dikelompokkan kedalam kelompok unit sehingga unit analisis dan kelompok unit merupakan suatu sistem hirarki ataumerupakan suatu struktur tersarang (Westet al., 2007). Dalam beberapa tahun terakhir, denganmengasumsikan bahwa data terobservasi adalah independen, model persamaan struktural telah diaplikasikan secara luasdalam penelitian pendidikan. Akan tetapi, masih terdapat penelitian fundamental dalam evaluasi pendidikan yang membutuhkan analisis data multilevel dari desain sampling yang berhirarki.

Untuk menganalisis pengaruh ini berdasarkan data multilevel diperlukan pemodelan statistik dari hubungan kausatif dan korelatif untuk masing-masing level. Beberapa tahun terakhir, kontribusi yang signifkan telah dikembangkan untuk analisis data multilevel dalam evaluasi pendidikan seperti dalam Goldstein (1987). Tujuan penelitian ini adalah untuk mempelajari konsep 2-level SEM dalam hal ini menggunakan metode algoritma EM sebagai alat estimasi parameter. Selanjutnya hasil dari penelitian ini diharapkan menambah wawasan bagi siapa saja, terutama yang mendalami bidang SEM sehingga dapat digunakan sebagai batu pijakan untuk penelitian yang lebih lanjut.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah studi literatur, yaitu mempelajari buku-buku dan jurnal-jurnal yang berkaitan dengan model persamaan struktural 2-level. Penelitian ini secara umum akan dibagi menjadi tiga tahap.

a. Menyusun rumusan model persamaan struktural 2 level. Untuk menyusun rumusan model tersebut, sebelumnya perlu dipelajari mengenai konsep-konsep yang lebih mendasar yaitu model persamaan struktural umum, pemodelan multilevel, komputasi statistik dan pemrograman dengan LISREL. Dimana konsep dasar tersebut akan digunakan dalam mempelajari lebih lanjut algoritma EM dalam melakukan estimasi parameter.
b. Reformulasi model persamaan struktural 2-level dengan algoritma EM sehingga diperoleh formulasi estimasi parameter serta uji fit dari model. Adapun langkah-langkah yang dilakukan yaitu: mempelajari formulasi umum dari model persamaan struktural dan konsep pemodelan multilevel. Selanjutnya mempelajari konsep dasar reformulasi model SEM 2 level dengan algoritma EM dan pengaplikasian formulasi tersebut dengan bantuan software LISREL. Setelah diperoleh formulasi estimasi, dilakukan simulasi komputasi dari data menggunakan formulasi tersebut.
c. Menerapkan formulasi dengan algoritma EM pada studi kasus. Studi kasus pada penelitian ini menggunakan data sekunder yang diperoleh dari Badan Standar Nasional Pendidikan (BSNP) dan Badan Pusat Statistik.

3. PEMBAHASAN

a. Model dan Asumsi

Diasumsikan x_{gi} adalah vektor random yang merupakan data yang diambil dari suatugrup ke-g padapengamatan ke-i. x_{gi} merupakan vektor berukuran $p \times 1$ dengan p adalahbanyaknya variabel atauindikator yang diambil pada setiap objek pengamatan. Modelstruktural 2-level untuk vektor ini sepertidalam Lee and Poon (1998) adalah

$$x_{gi} = v_g + v_{gi} \tag{1}$$

dengan $g = 1,2, \dots, G$ dan $i = 1,2, \dots, N_g$. G menyatakan banyaknya grup yang diamati dan N_g menyatakan banyaknya pengamatan tiap grup, dengan N_g tidak harus sama untuk setiap grup. v_g adalah vektor laten sebagai pembeda tiap grup dan v_{gi}vektor laten yang membedakan tiap pengamatan. Beberapa asumsi yang dibutuhkan untuk mengembangkan model ini adalah:

1. Vektor laten random level grup $\{v_g, g = 1,2, \dots, G\}$ berdistribusi normal identik dan independen, $v_g \sim N(0, \Sigma_B)$ dengan Σ_B adalah matriks kovariansi antar grup (*Between group*) dan $\Sigma_B > 0$.
2. Vektor laten random level pengamatan $\{v_{gi}, i = 1,2, \dots, N_g\}$ untuk suatu g berdistribusi identik dan independen, $v_{gi} \sim N(0, \Sigma_W)$ dengan Σ_W adalah matriks kovariansi dalam grup (*Within group*).
3. v_g dan v_{gi} saling independen.

Matriks kovariansi Σ_B dan Σ_W merupakan matriks atas vektor parameter $\theta_{q\times 1}$, dengan

$$\Sigma_B = \Sigma_B(\theta) \text{ dan } \Sigma_W = \Sigma_W(\theta) \text{ untuk setiap } g. \tag{2}$$

b. Estimasi Maksimum *Likelihood*

Estimasi yang digunakan pada model ini adalah maksimum *likelihood* mengingat distribusi yang diasumsikan adalah normal. Untuk mencari estimator maksimum *likelihood*, pertama-tama akan dicari fungsi bersama dari data terobservasi x_{gi} yang selanjutnya dimaksimumkan. Untuk membentuk fungsi bersama dimisalkan,

$$z_g = (x_{g1}, x_{g2}, \dots, x_{gN_g})' \text{ untuk setiap g = 1, 2, ..., G} \tag{3}$$

maka,

$$cov(z_g) = I_{N_g} \otimes \Sigma_W + J_{N_g} \otimes \Sigma_B, \tag{4}$$

dengan I_{N_g} adalah matriks identitas berordo $N_g \times N_g$ dan J_{N_g} adalah matriks yang setiap entrinya adalah satu berordo $N_g \times N_g$. Sehingga dengan menggunakan mean dan variansi pada Persamaan4 diperoleh $z_g = N[0, I_{N_g} \otimes \Sigma_W + J_{N_g} \otimes \Sigma_B]$. Jika $\Sigma_Z = I_{N_g} \otimes \Sigma_W + J_{N_g} \otimes \Sigma_B$ maka selanjutnya fungsi *likelihood* dari keseluruhan data terobservasi $z_1, \dots, z_g, \dots, z_G$adalah,

$$L(\theta|z_1, z_2, \dots, z_G) = (2\pi)^{-\sum_{g=1}^{G}\frac{N_g}{2}} |\Sigma_z|^{-\frac{\sum_{g=1}^{G}}{2}} \exp(\tfrac{1}{2} z_g' \Sigma_z^{-1} z_g) \tag{5}$$

selanjutnya persamaan 5 dilogaritma menjadi,

$$-2l(\theta) = N \ln(2\pi) + \textstyle\sum_{g=1}^{G} \left\{ \ln \left| \Sigma_z + z_g' \Sigma_z^{-1} z_g \right| \right\}. \tag{6}$$

Estimator $\hat{\theta}$ yang memaksimumkan persamaan 5 ekuivalen dengan meminimumkan fungsi pada persamaan 6. Karena suku pertama pada persamaan 6 merupakan suatu konstanta maka $\hat{\theta}$ meminimumkan fungsi,

$$F(\theta) = \sum_{g=1}^{G}\left\{\ln\left|\Sigma_z + z_g'\Sigma_z^{-1}z_g\right|\right\}. \tag{7}$$

Mengingat bahwa $\Sigma_Z = I_{N_g} \otimes \Sigma_W + J_{N_g} \otimes \Sigma_B$ maka dengan menggunakan sifat matriks $aI + bJ$ diperoleh nilai determinan dan invers dari matriks blok Σ_z yaitu,

$$|\Sigma_z| = |\Sigma_W|^{N_g-1}\left|\Sigma_W + N_g\Sigma_B\right|, \text{ dan} \tag{8}$$

$$\Sigma_z^{-1} = \left(I_{N_g} \otimes \Sigma_W^{-1}\right) - \left(J_{N_g} \otimes N_g^{-1}\left(\Sigma_W^{-1} - \Sigma_g^{-1}\right)\right) \text{ dengan } \Sigma_{\mathrm{g}} = \Sigma_W + N_g\Sigma_B. \tag{9}$$

Sehingga persamaan 7 dapat dituliskan sebagai,

$$\begin{aligned} F(\theta) = \sum_{g=1}^{G}\{&(N_g - 1)\ln|\Sigma_W|\} + \ln|\Sigma_W + N_g\Sigma_B| \\ &+ N_g^{-1}tr[\Sigma_g^{-1}\sum_{i,j}^{G}x_{gi}x_{gj}'] + N_g^{-1}tr[\Sigma_W^{-1}\sum_{i\neq j}^{G}(x_{gi}x_{gi}' - x_{gi}x_{gj}')]\}. \end{aligned} \tag{10}$$

Minimasi langsung pada persamaan 10 untuk menghasilkan estimator maksimum *likelihood*θ akan sangat rumit dan panjang meskipun Σ_B dan Σ_W berasal dari model yang sederhana seperti analisis faktor. Selanjutnya akan digunakan estimasi maksimum *likelihood* dengan menggunakan algoritma Ekspektasi dan Maksimasi (EM).

c. Estimasi Model Menggunakan Algoritma EM

Dinotasikan,

$$y_g = \binom{v_g}{z_g} \tag{11}$$

dengan $z_g$adalah variabel terobservasi pada grup ke-g seperti persamaan 3 dan v_g adalahvektor laten level-2. Pada algoritma EM, $y_g$dianggap sebagai himpunan data lengkap, $z_g$sebagai data tak lengkap, dan v_g sebagai missing value. Selanjutnya disimbolkan $X = (X_o, X_m)$untuk *Xo* data terobservasi, *Xm* dan *V* sebagai data hilang , yang akan digunakan untuk menghitung $L(X, V|\theta)$.

Mengingat bahwa $\{v_{gi}, i = 1, \dots, N_g\} \sim N_p(0, \Sigma_W)$ i.i.d pada grup yang sama dan $\{v_g, g = 1, \dots, G\} \sim N_p(0, \Sigma_B)$ i.i.d serta v_{gi} dan $v_g$tidak berkorelasi, maka untuk $i = 1,2, \dots, N_g$ dan $g = 1,2, \dots, G$ diperoleh,

$$-\log(\theta)=\frac{1}{2}\sum_{g=1}^{G}\log|\Sigma_B|+v_g'\Sigma_B^{-1}v_g+\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{N_g}\left(\log|\Sigma_W|+v_{gi}'\Sigma_W^{-1}v_{gi}\right). \quad (12)$$

Vektor v_{gi}pada persamaan 12 disubstitusikan dengan $x_{gi}-v_g$ sesuai dengan persamaan 1 sehingga diperoleh,

$$L(X,V|\theta)=\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{N_g}[\log|\Sigma_W|+\left(x_{gi}-v_g\right)'\Sigma_W^{-1}\left(x_{gi}-v_g\right)]$$
$$+\frac{1}{2}\sum_{g-1}^{G}[\log|\Sigma_B|+v_g'\Sigma_B^{-1}v_g].$$

Pada perhitugan langkah ekspektasi (E-step) pada algoritma EM, diperlukan nilai dari $E[L(X,V|\theta)|X,\theta]$, yaitu nilai ekspektasi negatif log-*likelihood* atas data lengkap bersyarat X dan θ berturut-turut sebagai data terobservasi dan parameter.

1) Langkah Ekspektasi (E-step)

Penentuan estimator θ dalam algoritma EM dilakukan secara iteratif, dan setiap iterasinya melibatkan dua langkah utama: langkah ekspektasi (E-step) dan langkah Maksimisasi (M-step). Pada iterasi ke-*r* langkah ekspektasi, perlu dicari ekspektasi bersyarat $L(X,V|\theta)$ jika diketahui X_o dan suatu nilai $\theta^{(r)}$awal. Sehingga didapatkan,

$$Q\left(\theta\middle|\theta^{(r)}\right)=E[L(X,V|\theta)|X_o,\theta=\theta^{(r)}] \quad (13)$$

Diperoleh hasil ekspektasi yang merupakan fungsi dalam θ. Selanjutnya disimbolkan,

$$C_g\left(\theta^{(r)}\right)=\frac{1}{N_g}\sum_{i=1}^{N_g}E[\left(x_{gi}-v_g\right)\left(x_{gi}-v_g\right)'|X_o,\theta^{(r)}] \text{ untuk } g=1,2,\ldots,G \quad (14)$$

dan

$$C_B\left(\theta^{(r)}\right)=E\left[v_gv_g'\middle|X_o,\theta^{(r)}\right]. \quad (15)$$

Persamaan 14 dan Persamaan 15 disubstitusikan pada Persamaan 13, sehingga diperoleh

$$Q\left(\theta\middle|\theta^{(r)}\right)=\sum_{g=1}^{G}\frac{N_g}{2}\left\{\log|\Sigma_W|+tr\left(\Sigma_W^{-1}C_g\left(\theta^{(r)}\right)\right)\right\}+\frac{G}{2}\left\{\log|\Sigma_B|+tr\left(\Sigma_B^{-1}C_B\left(\theta^{(r)}\right)\right)\right\}. \quad (16)$$

Ekspektasi bersyarat $X_o$dalam Persamaan 14 dan 15 cukup rumit. Nilai $C_g\left(\theta^{(r)}\right)$ dan $C_B\left(\theta^{(r)}\right)$ sangat panjang untuk diturunkan terlebih banyak data hilang yang dilibatkan. Maka dari itu, untuk nilai pendekatan $\hat{C}_g\left(\theta^{(r)}\right)$ atas $C_g\left(\theta^{(r)}\right)$ dan $\hat{C}_B\left(\theta^{(r)}\right)$ atas $C_B\left(\theta^{(r)}\right)$ harus ditemukan untuk mencari nilai pendekatan dari $\hat{Q}\left(\theta\middle|\theta^{(r)}\right)$ atas $Q\left(\theta\middle|\theta^{(r)}\right)$. Ide utamanya adalah mendekati nilai matriks $C_g\left(\theta^{(r)}\right)$ dan $C_B\left(\theta^{(r)}\right)$dengan observasi yang disimulasikan dari distribusi bersyarat yang sesuai, sehingga secara langsung dapat disimulasikan dari distribusi normal multivariat. Poon and Wang (2010) mengatakan hasil algoritma tersebut efisien, mudah diaplikasikan dan dapat dipraktikkan oleh para praktisi.

Selanjutnya, seluruh pengamatan pada suatu grup g dan variabel laten level grup dituliskan dalam $y_g$seperti pada Persamaan 11

$$y_g = \left(v_g, x_{g1}, x_{g2}, \ldots, x_{gN_g}\right)' = \left(v_g, z_g\right)'. \tag{17}$$

Di awal telah ditetapkan bahwa desain tak seimbang, sehingga $N_g$dapat berbeda untuk setiap g, berakibat dimensi $y_g$berbeda menurut g yaitu $p\left(N_g + 1\right)$. Karena $v_g \sim N(0, \Sigma_B)$ dan $\Sigma_Z \sim N[0, I_{N_g} \otimes \Sigma_W\left(\theta^{(r)}\right) + J_{N_g} \otimes \Sigma_B\left(\theta^{(r)}\right)]$ maka

$$\Omega_g = var\left(y_g\right) = \begin{pmatrix} \Sigma_B\left(\theta^{(r)}\right) & 1'_{\mathrm{N_g}} \otimes \Sigma_B\left(\theta^{(r)}\right) \\ 1_{\mathrm{N_g}} \otimes \Sigma_B\left(\theta^{(r)}\right) & I_{N_g} \otimes \Sigma_W\left(\theta^{(r)}\right) + J_{N_g} \otimes \Sigma_B\left(\theta^{(r)}\right) \end{pmatrix}.$$

Sehingga diperoleh distribusi dari $y_g$adalah $N\ [0, \Omega_g\ ]$, dengan 1_{N_g} adalah vektor satu berukuran $N_g \times\ 1$, $J_g$matriks satu berukuran $N_g \times\ N_g$, dan $I_g$adalah matriks identitas berukuran N_g. Tanpa mengurangi keumuman, z_g pada Persamaan 17 dipartisi menjadi x_{gm} dan x_{go} menjadi,

$$\begin{pmatrix} v_g \\ x_{gm} \\ \cdots \\ x_{go} \end{pmatrix}, \tag{18}$$

dengan x_{gm} adalah koleksi data hilang dari z_g dan x_{go} adalah koleksi data terobservasi. Matriks partisi dari $\Omega_g$atas Persamaan 18 menjadi

$$\begin{pmatrix} \Omega_{g,mm}\left(\theta^{(r)}\right) & \Omega'_{g,om}\left(\theta^{(r)}\right) \\ \Omega_{g,om}\left(\theta^{(r)}\right) & \Omega_{g,oo}\left(\theta^{(r)}\right) \end{pmatrix}.$$

Hasil distribusi bersyarat,

$$\begin{pmatrix} v_g \\ x_{gm} \end{pmatrix} \mid x_{go} \sim N\left[\Omega'_{g,om}\Omega_{g,oo}^{-1}x_{go}, \Omega_{g,mm} - \Omega'_{g,om}\Omega_{g,oo}^{-1}\Omega_{g,om}\right]. \tag{19}$$

Dari distribusi bersyarat Persamaan 19 diatas, dapat dilakukan simulasi secara langsung deretan variabel random sebanyak T :

$$\begin{pmatrix} \hat{v}_g^{(1)} \\ \hat{x}_{gm}^{(1)} \end{pmatrix}, \begin{pmatrix} \hat{v}_g^{(2)} \\ \hat{x}_{gm}^{(2)} \end{pmatrix}, \dots, \begin{pmatrix} \hat{v}_g^{(T)} \\ \hat{x}_{gm}^{(T)} \end{pmatrix}. \tag{20}$$

Pembangkitan variabel normal multivariat ini dapat dilakukan dengan program statistik seperti R dengan *package* mvtnorm. Tahap berikutnya adalah menambahkan nilai terobservasi x_{go}kedalam barisan 20, sehingga terbentuk barisan lain yaitu

$$\begin{pmatrix} \hat{v}_g^{(1)} \\ \hat{x}_{gm}^{(1)} \\ \hat{x}_{go}^{(1)} \end{pmatrix}, \begin{pmatrix} \hat{v}_g^{(2)} \\ \hat{x}_{gm}^{(2)} \\ \hat{x}_{go}^{(2)} \end{pmatrix}, \dots, \begin{pmatrix} \hat{v}_g^{(T)} \\ \hat{x}_{gm}^{(T)} \\ \hat{x}_{go}^{(T)} \end{pmatrix}, \tag{21}$$

dengan $\hat{x}_{gm}^{(j)} = x_{go}$ untuk setiap $j = 1, \dots, T$. Diperoleh

$$\begin{pmatrix} \hat{v}_g^{(1)} \\ \hat{x}_{g1}^{(1)} \\ \vdots \\ \hat{x}_{gN_g}^{(1)} \end{pmatrix}, \begin{pmatrix} \hat{v}_g^{(2)} \\ \hat{x}_{g1}^{(2)} \\ \vdots \\ \hat{x}_{gN_g}^{(2)} \end{pmatrix}, \dots, \begin{pmatrix} \hat{v}_g^{(T)} \\ \hat{x}_{g1}^{(T)} \\ \vdots \\ \hat{x}_{gN_g}^{(T)} \end{pmatrix}$$

sebagai barisan observasi yang telah dibangkitkan. Saat T cukup besar, maka diperolehestimator yang konsisten atas $C_g(\theta^{(r)})$ dan $C_B(\theta^{(r)})$pada Persamaan 14 dan 15, yaitu:

$$\hat{C}_g(\theta^{(r)}) = \frac{1}{T}\sum_{j=1}^{T} \frac{1}{N_g}\sum_{i=1}^{N_g}[\left(\hat{x}_{gi}^{(j)} - \hat{v}_g^{(j)}\right)\left(\hat{x}_{gi}^{(j)} - \hat{v}_g^{(j)}\right)'] \text{ untuk } g=1,2, \dots, G \tag{22}$$

dan

$$\hat{C}_B(\theta^{(r)}) = \frac{1}{T}\sum_{j=1}^{T} \frac{1}{G}\sum_{g=1}^{G}\left(\hat{v}_g^{(j)}\hat{v}_g^{(j)\prime}\right). \tag{23}$$

Ketika tidak ada observasi yang hilang dalam suatu grup g maka distribusi bersyarat 20 hanya berdimensi p dan barisan yang dibangkitkan seperti pada 21 hanya akan melibatkan $\hat{v}_g$.

2) Langkah Minimisasi (M-step)

Setelah $C_g(\theta^{(r)}), g = 1, \dots, G$ dan $C_B(\theta^{(r)})$diperoleh pada tahap ekspektasi, iterasi sebanyak r diperlukan dalam langkah minimisasi untuk meminimumkan fungsi dalam θ,

$$\hat{Q}\left(\theta\middle|\theta^{(r)}\right) = \sum_{g=1}^{G} \frac{N_g}{2}\left\{\log|\Sigma_W| + tr\left(\Sigma_W^{-1}\hat{C}_g\left(\theta^{(r)}\right)\right)\right\} + \frac{G}{2}\left\{\log|\Sigma_B| + tr\left(\Sigma_B^{-1}\hat{C}_B\left(\theta^{(r)}\right)\right)\right\}. \quad (24)$$

Fungsi ini sama seperti fungsi pencocokan (fit function) maksimum *likelihood* dalam Joreskog et al. (2000) yang terdapat pada program SEM standar seperti LISREL yaitu,

$$F(\theta) = \sum_{g=1}^{G} \frac{N_g}{N} F_g(\theta),$$

dengan $F_g(\theta) = F(\hat{z}_g, S_g, \mu_g(\theta), \Sigma_g(\theta))$ sebagai fungsi pencocokan untuk tiap grup, $\hat{z}_g$ dan S_g adalah mean vektor sampel dan matriks kovariansi pada grup g.

Sebagai catatan bahwa lebih rumit untuk memimumkan fungsi $F(\theta)$ pada 7 daripada $\hat{Q}\left(\theta\middle|\theta^{(r)}\right)$. Dari Persamaan 7 dapat dilihat bahwa setiap sukunya bergantung pada Σ_B dalam Σ_z. Sehingga meskipun Σ_W dan Σ_B memuat paramater yang sama ataupun berbeda, minimisasi atas $F(\theta)$ memperhatikan setiap parameternya. Hasilnya, usaha meminimasi $F(\theta)$ secara langsung dengan *scoring* atau algoritma yang lain menjadi menjadi lebih rumit daripada minimisasi $\hat{Q}\left(\theta\middle|\theta^{(r)}\right)$ menggunakan algoritma EM (Lee and Poon, 1998).

a) Penilaian Model Fit

Selanjutnya dapat dibentuk statistik uji goodness of fit (GOF) untuk menilai kecocokan model $\Sigma_W = \Sigma_W(\theta)$ dan $\Sigma_B = \Sigma_B(\theta)$. Diberikan θ_σ sebagai vektor parameter yang menghasilkan elemen-elemen segitiga bawah dari matriks Σ_W dan Σ_B saat matriks ini tanpa struktur (model dibawah H_0). Algoritma EM yang digunakan juga dapat digunakan untuk menentukan $\hat{\theta}_\sigma$atas θ_σ. Lebih spesifiknya, karena tidak ada struktur yang dikenakan pada Σ_W dan Σ_B, estimator yang meminimumkan fungsi 24 dapat dihasilkan sebagai elemen pada $\hat{C}_g\left(\theta^{(r)}\right)$dan $\hat{C}_B\left(\theta^{(r)}\right)$, dan tidak ada iterasi yang dibutuhkan pada M-step.

Setelah nilai $\hat{\theta}_\sigma$dihasilkan, uji statistik rasio *likelihood* dapat digunakan untuk menilai kecocokan model. Uji rasio *likelihood* sebagai berikut:

$$\lambda = \frac{F\left(X_o\middle|\widehat{\theta_\sigma}\right)}{F\left(X_o\middle|\hat{\theta}\right)}.$$

Sehingga *Goodness of fit* (GOF) adalah $-2\log\lambda$ yaitu,

$$GOF = -2\left(L\left(X_o\middle|\widehat{\theta_\sigma}\right) - L\left(X_o\middle|\hat{\theta}\right)\right)$$

dengan $L(X_o|\theta)$ adalah log $likelihood$dari data teramati, yaitu

$$L(X_o|\theta) = \sum_{g=1}^{G}\left(\log\left|\Omega_{g,oo}(\theta)\right| + x'_{go}\Omega_{g,oo}^{-1}(\theta)x_{go}\right).$$

3) STUDI KASUS

a) Data

Populasi dari studi kasus ini adalah seluruh sekolah menengah atas (SMA) di Indonesia baik negeri maupun swasta. Adapun sampel yang diambil adalah sekolah-sekolah di 109 kabupaten/kota yang tersebar di pulau-pulau terbesar di Indonesia. Beberapa sampel yang diambil adalah kabupaten di Provinsi Jawa Tengah sebanyak 35 kabupaten/kota, Sumatra Barat sebanyak 19 Kabupaten, Kalimantan Tengah sebanyak 14 Kabupaten, Sulawesi Tengah 11 Kabupaten, Papua 29 Kabupaten, Nusa Tenggara Barat 10 Kabupaten/kota. Total sekolah yang dijadikan sampel sebanyak 1524 sekolah. Data merupakan data sekunder yaitu "LAPORAN HASIL UJIAN NASIONAL 2012/2013", sebuah apikasi yang diterbitkan oleh Badan Standar Nasional Pendidikan (BSNP). Selanjutnya dari data tersebut dilakukan analisis faktor tunggal 2-level (*single factor analysis*) yang estimasinya menggunakan ML (maksimum *likelihood*) dengan algoritma EM dengan bantuan *software* R 3.0.3 dan LISREL 8.8 *Student Edition*.

b) Model Pengukuran

Model pengukuran untuk model baik beetwen maupun within adalah

- matematika= λ_1 Kualitas Pendidikan Sains + δ_1
- fisika= λ_2 Kualitas Pendidikan Sains + δ_2
- kimia= λ_3 Kualitas Pendidikan Sains + δ_3
- biologi= λ_4 Kualitas Pendidikan Sains + δ_4

dengan λ adalah koefisien regresi model pengukuran, δ adalah error variabel indikator dari variabel laten.

Model pengukuran baik beetwen-grup maupun within-grup terdiri dari empat buah variabel *observed* (indikator), sehingga p=4. Banyaknya parameter yang akan diestimasi adalah 8, yaitu 4 buah *loading* faktor (λ) dan 4 buah error varians (δ). Jadi derajat bebas model between maupun within adalah $\left(\frac{1}{2}\times 4\times 5\right) = 2$. Dengan derajat bebas yang positif maka syarat perlu sebagai model yang *identified* telah terpenuhi.

c) Uji Validitas Konstruk

Model Between

- Signifikansi parameter dan *loading* faktor (λ)

Dapat dilihat pada Tabel 1 bahwa nilai estimator loading faktor untuk modelbetween dari $b\lambda_1, b\lambda_2, b\lambda_3$ dan $b\lambda_4$berturut-turut adalah 0.896, 0.671, 1.053, 0.780.Keempat loading faktor ini mempunyai nilai t-value yang besar yaitu lebih besardari 1,96 (α= 5%), sehingga nilai parameter tersebut dinyatakan signifikan.

Tabel 1: hasil estimasi parameter grup between

parameter	estimator	nilai terstndarisasi	Standar error	t-*value*
bλ_1	0.896	0.897	0.078	11.44
bλ_2	0.671	0.605	0.076	8.845
bλ_3	1.053	0.945	0.079	13.252
bλ_4	0.780	0.902	0.067	11.713
bδ_1	0.135	0.135	0.028	4.799
bδ_2	0.240	0.195	0.043	5.540
bδ_3	-0.012	-0.009	0.021	-0.571
bδ_4	0.082	0.110	0.020	4.196

- Reliabilitas indikator

Nilai reliabilitas indikator (R^2) dapat dilihat dari output berikut ini.

Tabel 2: Nilai R^2 indikator grup between

Matematika	Fisika	Kimia	Biologi
0.856	0.652	1.011	0.881

Hasil output menunjukkan nilai reliabilitas untuk keempat indikator lebih besar dari 0.5 sehingga keempat indikator dinyatakan reliabel.

- Reliabilitas Konstruk

Nilai loading terstandarisasi pada tabel 4.1 digunakan untuk menghitung koefisien reliabilitas komposit (CR) atau sering juga disebut reliabilitas konstruk, dan *average variance extracted* (AVE). Hasil perhitungan CR dan AVE menghasilkan nilai CR sebesar 0.91 dan nilai AVE yang dihasilkan adalah 0.72. Nilai CR dan AVE minimal yang disarankan adalah 0.7 dan 0.5. Jadi menggunakan ukuran CR dan AVE menunjukkan bahwa reliabilitas konstruk pada model-between ini baik.

Model Within

- Signifikansi parameter dan loading faktor (λ)

Dapat dilihat pada Tabel 3 bahwa estimator loading faktor untuk model within berturut-turut dari $w\lambda_1, w\lambda_2, w\lambda_3$ dan $w\lambda_4$ adalah 0.782, 0.782, 0.808, dan 0.609. Jika dilihat dari nilai t-value keempat indikator pada model ini maka keempatnya signifikan karena semua nilai t-value bernilai lebih besar dari 1.96 (α = 5%).

Tabel 3: hasil estimasi parameter grup within

parameter	estimator	nilai terstndarisasi	Standar error	t-*value*
wλ_1	0.782	0.783	0.025	31.601
wλ_2	0.782	0.703	0.028	27.617
wλ_3	0.808	0.725	0.028	28.710
wλ_4	0.609	0.704	0.022	27.631
wδ_1	0.387	0.388	0.023	17.098
wδ_2	0.624	0.505	0.030	20.751
wδ_3	0.589	0.474	0.029	19.999
wδ_4	0.378	0.505	0.018	20.766

- Reliabilitas indikator

Nilai reliabilitas indikator (R^2) untuk model-within ada pada output berikut ini.

Tabel 4: Nilai R^2 indikator grup between

Matematika	Fisika	Kimia	Biologi
0.612	0.495	0.526	0.495

Nilai reliabilitas R^2 atas keempat indikator menunjukkan nilai yang besar yaitu diatas 0.5. Dapat disimpulkan bahwa keempat indikator ini reliabel pada model *within*.

- Reliabilitas Konstruk

Nilai CR dan AVE untuk grup within hasil perolehan dari perhitungan estimasi parameter terstandarisasi pada tabel 3 adalah 0.82 dan 0.53. Hasil ini sudah melebihi ukuran CR dan AVE minimal yang disarankan. Dapat disimpulkan bahwa reliabilitas konstruk pada model-within ini baik dan indikator-indikator tersebut representatif bagi konstruknya.

d) Penilaian Model Fit

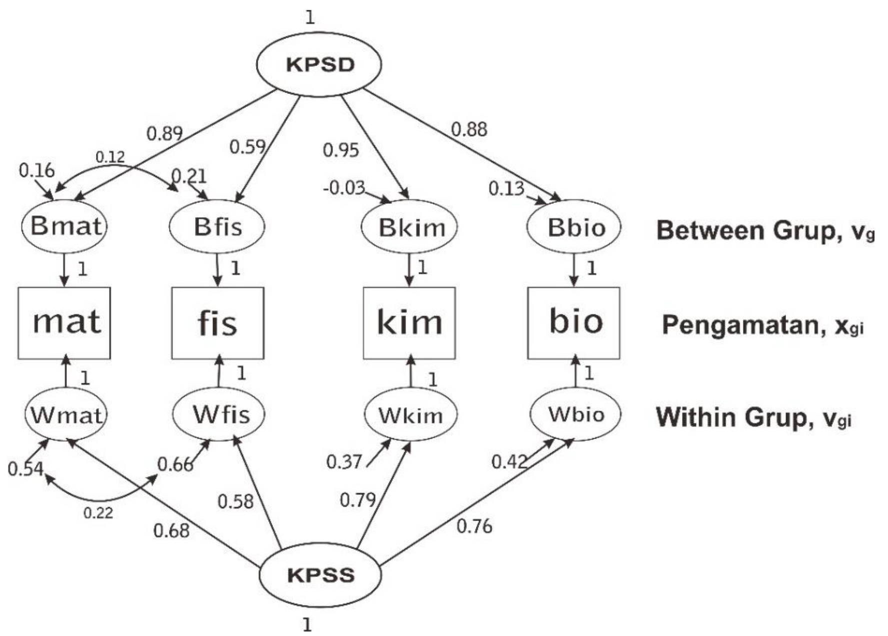
Ukuran tingkat kecocokan keseluruhan model dapat dilihat pada output programlisrel berikut ini.

```
Global Goodness of Fit Statistics, Multilevel SEM
-2ln(L) for the saturated model =       16018.763
-2ln(L) for the fitted model    =       16182.083
Degrees of Freedom = 4
Full Information ML Chi-Square  = 163.321 (P = 0.0)
Root Mean Square Error of Approximation (RMSEA) = 0.228
90 Percent Confidence Interval for RMSEA = (0.199 ; 0.258)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.000
```

Statistik chi-kuadrat (χ^2) yang dihasilkan sebesar 163.321 dengan pvalue 0.00 (<0.05). Nilai chi-kuadrat ini cukup besar ditambah nilai p-value yang sangat kecil mengindikasikan bahwa kecocokan model kurang baik. Ukuran berikutnya adalah *Root Mean Square Error of Approximation* (RMSEA) yang bernilai 0.228 (*poor fit*). 90% interval konfidensi dari RMSEA = (0.199 ; 0.258) menandakan bahwa estimasi RMSEA mempunyai presisi yang baik. P-Value for Test of Close Fit (RMSEA < 0,05) = 0.000 berarti kecocokan model kurang baik, karena untuk mendapat model yang baik nilai *p-value* adalah > 0,05.

e) Perbaikan Model

Berdasarkan uraian sebelumnya yang mengatakan bahwa model kurang baik berdasarkan nilai chi-kuadrat dan RMSEA maka diperlukan respesifikasi terhadap model tersebut. Perbaikan model ini didasarkan atas *modification index* yang terdapat pada luaran program dan disiplin ilmu pendidikan. Modifikasi yang dilakukan adalah dengan memberikan korelasi antara indikator matematika dan fisika sebagaimana penelitian (Kereh et al., 2014) yang menyebutkan adanya korelasi tinggi antara matematika dasar dengan fisika. Selanjutnya menggunakan program yang sama diperoleh hasil yang dapat dilihat pada gambar berikut.



Gambar 1: Diagram lintasan 2 level analisis faktor dengan korelasi indikator fisika dan matematika.

Respesifikasi model diatas memperbaiki nilai chi-kuadrat menjadi 2.69 dengan pvalue =0.26, dan menurunkan nilai RMSEA dari 0.228 menjadi 0.021. Dapat disimpulkan: pertama, untuk nilai chi-kuadrat 2.69

dengan p-value =0.26 menunjukkan bahwa model fit. Kedua, *P-Value for Test of Close Fit* (RMSEA < 0.05) = 0.92 menunjukkan bahwa nilai RMSEA=0.021 signifikan sehingga disimpulkan model *close fit*. Model yang telah dimodifikasi inilah yang selanjutnya dilakukan pembahasan.

Ukuran kecocokan model setelah dimodifikasi adalah sebagai berikut,

```
            Global Goodness of Fit Statistics, Multilevel SEM
            -2ln(L) for the saturated model =       16018.763
            -2ln(L) for the fitted model    =       16021.453
                              Degrees of Freedom = 2
            Full Information ML Chi-Square  = 2.69 (P = 0.26)
         Root Mean Square Error of Approximation (RMSEA) = 0.021
        90 Percent Confidence Interval for RMSEA = (0.0 ; 0.078)
           P-Value for Test of Close Fit (RMSEA < 0.05) = 0.92
```

f) Pembahasan studi kasus

Setelah model dimodifikasi maka tahap terakhir adalah interpretasi dari model tersebut yang telah dinyatakan baik tingkat kecocokannya. Untuk model within, model pengukuran terstandarisasi adalah,

$$\text{Matematika} = 0{,}68 * \text{KPSS} + 0{,}54$$
$$\text{Fisika} = 0{,}58 * \text{KPSS} + 0{,}66$$
$$\text{Kimia} = 0{,}79 * \text{KPSS} + 0{,}37$$
$$\text{Biologi} = 0{,}76 * \text{KPSS} + 0{,}42.$$

Dapat disimpulkan bahwa KPS tingkat sekolah memberikan pengaruh sangat besar terhadap pencapaian nilai UN Matematika, Fisika, Kimia, Biologi. Hasil perhitungan terstandarisir pada Persamaan 25 menyebutkan KPS berpengaruh $0{,}68^2$ (46%) untuk Matematika, Fisika $0{,}58^2$ (34%), Kimia $0{,}79^2$ (62%) dan Biologi $0{,}76^2$ (58%). Hal ini menunjukkan, bahwa dengan mengontrol perbedaan antar kabupaten, KPS tingkat sekolah memprediksikan pencapaian rata-rata nilai UN untuk keempat mapel tersebut dengan signifikan.

Ketiga, hal serupa juga berlaku untuk faktor KPS tingkat kabupaten yang berpengaruh (signifikan pada taraf 5%) terhadap pencapaian rata-rata UN sekolah di mapel rumpun sains tersebut. Dapat dilihat pada hasil model pengukuran terstandar berikut.

$$\text{Matematika} = 0{,}89 * \text{KPSD} + 0{,}16$$
$$\text{Fisika} = 0{,}59 * \text{KPSD} + 0{,}21$$
$$\text{Kimia} = 0{,}95 * \text{KPSD} - 0{,}03$$
$$\text{Biologi} = 0{,}88 * \text{KPSD} + 0{,}13.$$

Hasil tersebut menunjukkan pengaruh KPS tingkat daerah terhadap capaian UN sainssekolah, Matematika 79%, Fisika 35%, Kimia 90%, dan Biologi 77%.

4. KESIMPULAN

Beberapa kesimpulan yang dapat diambil dari penelitian ini adalah:

a. Model Persamaan struktural 2-level dapat digunakan untuk menganalisis model Persamaan struktural untuk struktur data yang berhirarki.
b. Penggunaan algoritma ekspektasi-maksimasi (EM) diperlukan untuk menghasilkan estimator data hilang sehingga dapat menggunakan metode maksimum *likelihood*.
c. Dari prosedur analisis yang digunakan diperoleh estimator, statistik chi-kuadrat (χ^2) dan ukuran kesalahan RMSEA untuk model. Nilai tersebut selanjutnya digunakan untuk mengevaluasi model fit ataukah tidak.
d. Analisis studi kasus menghasilkan model yang fit baik untuk model level-2 (modelbetween) maupun level-1 (model-within). Yaitu keempat indikator (Matematika, Fisika, Kimia, Biologi) merupakan variabel representasi atas variabel laten KPS (Kualitas Pembelajaran Sains).

5. DAFTAR PUSTAKA

Bollen, K. (1989). *Structural Equation with latent Variables*. John Wiley & Sons.

Ghozali, I. (2008). SEM: *Konsep dan Aplikasi dengan program LISREL 8.8.*Semarang : UNDIP Press.

Goldstein, H. (1987). *Multilevel Model in Educational and Social Research.*Newyork : Oxford University Press.

Joreskog, K. G., Sorbom, D., du Toit, S., and du Toit, M. (2000). *LISREL 8: New Statistical Features*. Scientific Software International, Inc.

Kereh, C., Liliasari, P.C.Tjiang, and J.Sabandar (2014). Korelasi penguasaan materi matematika dasar dengan penguasaan materi pendahuluan fisika inti. *Jurnal Pendidikan Fisika Indonesia*, 10(2):140-149.

Lee, S. Y. and Poon, W.Y. (1998). Analysis of two-level structural equation models via EM type algorithm. *Statistica Sinica*, 8:749-766.

Poon, W.Y. and Wang, H.B. (2010). Analysis of a two-level structural equation model with missing data. *Sociological Methods & Research*: SAGE, 39(1):25-55.

West, B., Welch, K., and Galechi, A. (2007). Linear Mixed Models: *A Practical Guide Using Statistical Software*. Chapman n Hall. Boca Raton