

PENGARUH UKURAN SAMPEL DAN INTRACLASS CORRELATION COEFFICIENTS (ICC) TERHADAP BIAS ESTIMASI PARAMETER MULTILEVEL LATENT VARIABLE MODELING: STUDI DENGAN SIMULASI MONTE CARLO

Muhammad Dwirifqi Kharisma Putra^{1*}, Jahja Umar¹, Babrul Hayat¹, Agung Priyo Utomo²

¹UIN Syarif Hidayatullah Jakarta, ²Sekolah Tinggi Ilmu Statistik

¹Ciputat, Cempaka Putih, Ciputat Timur, Tangerang Selatan, Banten 15412, Indonesia

²Kampung Melayu, Jatinegara, RT.1/RW.4, Jakarta Timur, DKI Jakarta 13330, Indonesia

* Corresponding Author: muhammad.dwirifqi@gmail.com

Abstrak

Studi ini menggunakan simulasi *Monte Carlo* dilakukan untuk melihat pengaruh ukuran sampel dan *intraclass correlation coefficients* (ICC) terhadap bias estimasi parameter *multilevel latent variable modeling*. Kondisi simulasi diciptakan dengan beberapa faktor yang ditetapkan yaitu lima kondisi ICC (0.05, 0.10, 0.15, 0.20, 0.25), jumlah kelompok (30, 50, 100 dan 150), jumlah observasi dalam kelompok (10, 20 dan 50) dan diestimasi menggunakan lima metode estimasi: ML, MLF, MLR, WLSMV dan BAYES. Jumlah kondisi keseluruhan sebanyak 300 kondisi dimana tiap kondisi direplikasi sebanyak 1000 kali dan dianalisis menggunakan software Mplus 7.4. Kriteria bias yang masih dapat diterima adalah $< 10\%$. Hasil penelitian ini menunjukkan bahwa bias yang terjadi dipengaruhi oleh ukuran sampel dan ICC, penelitian ini juga menunjukkan bahwa metode estimasi WLSMV dan BAYES berfungsi lebih baik pada berbagai kondisi dibandingkan dengan metode estimasi berbasis ML.

Kata kunci: *multilevel latent variable modeling, intraclass correlation coefficients, Metode Markov Chain Monte Carlo*

THE IMPACT OF SAMPLE SIZE AND INTRACLASS CORRELATION COEFFICIENTS (ICC) ON THE BIAS OF PARAMETER ESTIMATION IN MULTILEVEL LATENT VARIABLE MODELING: A MONTE CARLO STUDY

Abstract

A monte carlo study was conducted to investigate the effect of sample size and intraclass correlation coefficients (ICC) on the bias of parameter estimates in multilevel latent variable modeling. The design factors included (ICC: 0.05, 0.10, 0.15, 0.20, 0.25), number of groups in between level model (NG: 30, 50, 100 and 150), cluster size (CS: 10, 20 and 50) to be estimated with five different estimator: ML, MLF, MLR, WLSMV and BAYES. Factors were interegated into 300 conditions (4 NG \times 3 CS \times 5 ICC \times 5 Estimator). For each condition, replications with convergence problems were exclude until at least 1.000 replications were generated and analyzed using Mplus 7.4, we also consider absolute percent bias $< 10\%$ to represent an acceptable level of bias. We find that the degree of bias depends on sample size and ICC. We also show that WLSMV and BAYES estimator performed better than ML-based estimator across varying sample sizes and ICC's conditions.

Keywords: *multilevel latent variable modeling, intraclass correlation coefficients, Markov Chain Monte Carlo method*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v21i1.12895>

Pendahuluan

Dalam bidang penelitian sosial terkadang diperoleh struktur data yang merupakan data hirarki (*hierarchical*). Data yang terstruktur hirarki merupakan data yang timbul karena individu-individu terkumpul dalam kelompok-kelompoknya, dimana individu-individu dalam kelompok yang sama memiliki karakteristik yang cenderung sama. Struktur hirarki mengindikasikan bahwa data yang dianalisis berasal dari beberapa *level*, dimana *level* yang lebih rendah tersarang pada *level* yang lebih tinggi. Permodelan *multilevel* umumnya digunakan pada data yang berstruktur hirarki (*clustered, nested*) (de Leeuw & Meijer, 2008; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002). Ada dua masalah utama yang mungkin muncul akibat mengabaikan struktur hirarki. Pertama, jika analisis dilakukan dengan mengumpulkan data unit *level* rendah ke unit *level* tinggi (*aggregation*), maka banyak informasi yang hilang dan analisis statistika menjadi kehilangan kekuatan. Kedua, apabila seorang peneliti tidak hati-hati dalam menginterpretasi hasil maka akan menimbulkan kesalahan seperti menganalisis data pada salah satu *level* dan merumuskan kesimpulan pada *level* lain. (Kaplan, Kim & Kim, 2009; Muthén, 1994).

Metode analisis data pada umumnya menggunakan asumsi berupa penyederhanaan bahwa data yang telah diperoleh sebagai *simple random sample* dari suatu populasi tertentu. Ini melibatkan asumsi *independently and identically distributed observation (IID)*. Data pendidikan banyak yang, bagaimanapun, diperoleh melalui, desain sampel yang kompleks, seperti *multistage sampling* yang melibatkan pengamatan berbentuk *cluster* di mana asumsi IID tidak realistis. (Muthén, 1991). *Multistage sampling* digunakan ketika sampel diambil secara acak dari unit yang lebih tinggi dan disampel dari populasi yang lebih besar dari unit tersebut. Hasil pengambilan sampel menggunakan metode *multistage sampling* pada data hierarki yang terstruktur (misalnya, siswa berkumpul di dalam satu unit kelas), membuat residual bergantung pada variasi *between-cluster*. Skor pada variabel yang diamati dari anak-anak dalam satu

kelas mungkin lebih mirip daripada anak-anak di kelas yang berbeda, misalnya. Apabila mengabaikan struktur data hirarki dapat terjadi bias pada estimasi hubungan antar item (Geldhof, Preacher & Zyphur, 2014).

Selama bertahun-tahun, pemodelan *multilevel* telah mengalami perkembangan yang sangat pesat dan aplikasi untuk mengklaim bahwa pemodelan *multilevel* sekarang tegas berlindung di berbagai metodologi untuk penelitian ilmu sosial dan perilaku. Selain itu, saat ini telah banyak dilakukan untuk mengintegrasikan model bertingkat dengan *structural equation modeling (SEM)* sehingga memberikan metodologi umum yang dapat menjelaskan masalah kesalahan pengukuran, mediasi, dan simultanitas. Hal ini berguna, karena itu, untuk memberikan gambaran tentang metodologi, memeriksa perkembangan terbaru, dan menawarkan jalan untuk upaya penelitian masa depan (Kaplan, Kim & Kim, 2009).

Dalam berbagai kasus, penggunaan metode CFA yang di desain untuk data berstruktur *single-level* akan menyebabkan bias pada estimasi parameter dan *standard error* (Julian, 2001). MCFA yang merupakan submodel dari MSEM, telah dikembangkan untuk mengatasi permasalahan ini. (Muthén & Asparouhov, 2009). Tetapi teknik ini bukan merupakan teknik dengan sampel kecil. Secara khusus, teknik ini dapat digunakan untuk data yang memiliki jumlah kelompok yang cukup besar, setidaknya sekitar 50-100. Seperti yang kemukakan oleh Cronbach (Muthén, 1991), jika terbentur biaya, mungkin lebih baik mengamati siswa per kelas yang lebih sedikit untuk membantu menambah kelas yang lebih banyak. Sejalan dengan pendapat tersebut Hayes (2006) juga berpendapat bahwa permodelan *multilevel* merupakan prosedur dengan "sampel yang besar", yang berarti bahwa analisis matematika dan asumsi teoritis yang mendasari statistik yang dihasilkan oleh permodelan *multilevel* didasarkan pada analisis statistik yang dihitung dalam sampel yang besar. Seperti biasa, betapa besar cukup besar adalah pertanyaan yang sulit dijawab, karena tergantung pada banyak hal. Lebih parah lagi, ukuran sampel

yang dibutuhkan adalah fungsi dari kedua jumlah unit *level 1* dan jumlah unit *level 2*.

Dalam mendesain penelitian *multi-level*, peneliti harus memberi perhatian pada berbagai permasalahan terkait ukuran sampel dan bagaimana distribusi dari individu dalam unit yang akan diteliti (Heck & Thomas, 2015). Sejalan dengan pendapat tersebut, ukuran sampel yang cukup merupakan salah satu masalah yang paling penting dalam pemodelan *multilevel*, Kondisi desain yang paling dasar seperti sejumlah kelompok pada setiap tingkat analisis dan ukurannya menentukan kemampuan untuk memperoleh hasil estimasi koefisien regresi yang *unbiased* serta *standard errors* dengan besaran yang masih dapat diterima dan kekuatan tes (Łaszkiewicz, 2013; Snijders, 2005).

Selain itu, Busing (Łaszkiewicz, 2013) menemukan ukuran sampel yang tidak mencukupi (10 sampai 50 kelompok dengan 5 atau 10 orang) mungkin menjadi penyebab yang membuat model menjadi tidak kon-vergen. Meskipun sifat *asymptotic* dari estimator pada model *multilevel* (seperti REML atau IGLS), penggunaan ukuran sampel yang lebih besar menjamin pengurangan bias, di tengah ketertarikan mengetahui batas bawah dari sampel (Mass & Hox, 2005). Dengan demikian, ukuran sampel yang memadai (cukup) bisa dianggap sebagai sampel minimum, yang menjamin *unbiasedness* (atau lebih tepatnya: ukuran bias yang masih bisa diterima).

Perbedaan pada estimasi parameter dengan metode *single-level* dan dibandingkan dengan metode *clustering* (model *multilevel*) bergantung pada beberapa aspek dari data. (Pornprasertmanit, Lee & Preacher, 2014). Dengan pengukuran konstruk tingkat *cluster*, berbagai langkah-langkah pengukuran digunakan untuk mengevaluasi apakah respon dari item menunjukkan besaran dari *clustering* (pengelompokan) seperti yang diharapkan pada konstruk tingkat *cluster*. Salah satu ukuran yang sering digunakan adalah ICC. (Shrout & Fleis, 1979 dalam Stapleton, Yang & Hancock, 2016).

ICC merupakan salah satu hal yang harus diperhitungkan karena hal ini merupa-

kan hal penting yang dapat merubah varians error pada model regresi linier sederhana (Kreft & de Leeuw, 1998). ICC merupakan faktor penting yang menjelaskan variabilitas keseluruhan yang dijelaskan oleh unit *level 2* yang menjadi faktor penting dalam permodelan *multilevel* (Julian, 2001; Pornprasertmanit, Lee & Preacher, 2014). Terdapat keterkaitan antara rendahnya ICC dan besarnya bias pada estimasi parameter pada model *between-level* (Preacher, Zhang & Zyphur, 2011) serta rendahnya ICC dan kaitannya dengan rendahnya tingkat konvergensi (Kim, Kwok & Yoon, 2012 dalam Hsu et. al., 2016). Metode estimasi dengan estimator *maximum likelihood* juga tidak dapat bekerja dengan baik saat nilai ICC rendah ataupun jumlah individu pada *cluster* dibawah 50 orang. (Hox & Mass, 2004 dalam Stapleton, Yang & Hancock, 2016). ICC nilainya berkisar dari 0 ke 1, nilai ICC yang tinggi menunjukkan proporsi varians yang lebih besar dari varians pada tingkat “antar” dan bias sehingga kemungkinan besar jika sifat data bertingkat tersebut tidak diperhitungkan. (Dyer, Hanges & Hall, 2005).

Untuk menjawab permasalahan tersebut, baik permasalahan terkait ukuran sampel dan juga permasalahan terkait besarnya ICC dapat dilakukan studi dengan simulasi data yang memungkinkan peneliti menjawab pertanyaan-pertanyaan seputar masalah tersebut, salah satunya dengan menggunakan simulasi *Monte Carlo*. Terdapat beberapa penelitian yang menggunakan simulasi *Monte Carlo* pada analisis data *multilevel* seperti untuk menginvestigasi efek dari mengabaikan *clustering* dari data serta berfokus pada ICC dan faktor penting lain sebagai penyebabnya (Pornprasertmanit, Lee & Preacher, 2014),

Studi dengan simulasi data mulai populer pada akhir abad ke 19 dan awal abad ke 20 yang digunakan pada berbagai bidang ilmu pengetahuan. (Feinberg & Rubright, 2016). Terdapat berbagai metode dari simulasi data, salah satunya adalah simulasi *Monte Carlo* dengan algoritma yaitu *Markov Chain Monte Carlo* (MCMC) yang menggunakan pemilihan angka secara acak untuk menye-

lesaikan permasalahan *modeling* yang sulit ini, metode ini diperkenalkan pada bidang psikometri oleh Patz dan Junker tahun 1999 (Gelfand & Smith, 1990; Patz & Junker, 1999 dalam Feinberg & Rubright, 2016). Umumnya, studi menggunakan simulasi *Monte Carlo* yang digunakan pada bidang SEM untuk mempelajari sifat dari estimator dan uji statistik yang digunakan pada berbagai kondisi yang dimanipulasi oleh peneliti, seperti ukuran sampel, besarnya kesalahan spesifikasi model dan tidak normalnya data (Brown, 2006). Sehingga dari latar belakang yang telah dijelaskan sebelumnya, akan diuji pengaruh ukuran sampel dan *intraclass correlation coefficients* (ICC) terhadap bias estimasi parameter multilevel *latent variable modeling* dengan menggunakan metode berbasis pendekatan *Bayesian* yaitu *markov chain monte carlo* (MCMC).

Para peneliti telah lama menyadari permasalahan ini, dalam bidang pendidikan telah terjadi perdebatan yang dimulai mengenai permasalahan '*unit of analysis*' pada tingkat berbeda (Burstein, 1980). Sebelum permodelan *multilevel* berkembang dan menjadi metode penelitian yang umum digunakan, permasalahan yang diakibatkan karena mengabaikan struktur hirarki dari data masih dapat diterima dan dimengerti karena sulit untuk menyelesaikan masalah tersebut karena perangkat lunak yang belum tersedia (Goldstein, 2011). Tetapi pada saat ini telah banyak metode yang dapat digunakan untuk mengatasi permasalahan terkait metode analisis yang tepat untuk menganalisis data dengan struktur hirarki yang akan dijelaskan pada bagian selanjutnya.

Konsep Dasar Multilevel Latent Variable Modeling

Untuk mengatasi keterbatasan yang terkait dengan pendekatan bermasalah untuk analisis data yang berstruktur hirarki, Para peneliti telah membuat analisis untuk bentuk data hirarkis yang memungkinkan untuk pemodelan sesuai sistem organisasi seperti sekolah. Selain perkembangan statistik, kemajuan perangkat lunak sekarang memungkinkan estimasi relatif mudah untuk model

multilevel, dan pemodelan seperti sekarang umum dilakukan dalam ilmu-ilmu sosial dan perilaku (Kaplan, Kim & Kim, 2009). Terdapat pengembangan dari analisis *multiple regression* yang dikenal dengan *multilevel modeling* (MLM) yang juga dikenal dengan *hierarchical linear modeling*, *random coefficient modeling* atau *mixed effect modeling* yang merupakan metode statistik berbasis regresi yang digunakan apabila data terstruktur hirarki atau dalam *cluster* (Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002) sedangkan pengembangannya pada penggunaan variabel laten didalamnya yaitu *multilevel latent variable modeling* (MLLVM) yang perkembangannya dirangkum oleh berbagai literatur (misal, Kaplan, Kim & Kim, 2009). Pendekatan ini didasarkan pada metode SEM. Muthén & Satorra (1995) menjelaskan pendekatan SEM yang digunakan pada data yang berasal dari survei yang kompleks dapat berupa agregasi dan disagregasi. Pendekatan berbasis SEM ini juga dikenal dengan nama *covariance structure analysis* yang menggunakan model matematika untuk menjelaskan matriks kovarians dari sekumpulan variabel dari faktor yang jumlahnya lebih sedikit dan dapat diperpanjang untuk menganalisis struktur yang ada. (Heck & Thomas, 2015; Muthén, 1994). Pada penelitian ini *multilevel latent variable modeling* yang dimaksud, dibatasi pada penggunaan metode *multilevel* CFA yang diperkenalkan oleh Muthén (1991, 1994).

Konsep Dasar Multilevel CFA

Stapleton, Yang & Hancock (2016) menjelaskan bahwa sebuah permasalahan muncul pada CFA apabila data dikumpulkan dari individu dengan struktur hirarki ataupun *nested*. Pengukuran pada tingkat individu diharapkan responnya relevan dengan hipotesis dari konstruk dan oleh karena itu skor pada pengukuran tersebut harus mencerminkan variabilitas individu.

Kaplan, Kim & Kim (2009) menjelaskan bahwa untuk memulai pendekatan *Multilevel Confirmatory Factor Analysis* (MCFA), dimulai dari mempertimbangkan model yang terurai p-dimensi vektor respon y_{ig} untuk siswa i di sekolah g ke dalam *sum of a grand*

mean μ , bagian *between-group* v_g dan bagian *within-group* u_{ig} . Sebagai berikut:

$$y_{ig} = \mu + v_g + u_{ig} \dots (1)$$

Matriks kovarians untuk vektor respon dapat ditulis sebagai y_{ig} :

$$\Sigma_T = \Sigma_B + \Sigma_w \dots (2)$$

di mana Σ_T adalah matriks kovarians keseluruhan populasi, Σ_b adalah matriks kovarians populasi *between-group*, dan Σ_w adalah matriks kovarians populasi *within-group*. Jumlah sampel dapat didefinisikan sebagai:

$$\bar{y}_{.g} = \frac{1}{n_g} \sum_{i=1}^{n_g} \bar{y}_{ig} \dots (3)$$

$$\bar{y} = \frac{1}{N} \sum_{g=1}^G \sum_{i=1}^{n_g} \bar{y}_{ig} \dots (4)$$

$$S_w = \frac{1}{N - G} \sum_{g=1}^G \sum_{i=1}^{n_g} (\bar{y}_{ig} - \bar{y}_{.g})(\bar{y}_{ig} - \bar{y}_{.g})' \dots (5)$$

$$S_b = \frac{1}{G - 1} \sum_{g=1}^G n_g (\bar{y}_{.g} - \bar{y})(\bar{y}_{.g} - \bar{y})' \dots (6)$$

di mana $\bar{y}_{.g}$ adalah *mean* sampel untuk kelompok g , \bar{y} adalah *grand mean*, S_w adalah matriks kovarians *sample pooled within group*, dan S_b adalah matriks kovarians *between groups*. Seperti aplikasi standar regresi linier pada data yang didapat dari pengambilan sampel secara *multistage*, penerapan analisis faktor juga harus memperhitungkan efek bersarang (*nested*). Misalnya, sekumpulan item skala sikap digunakan untuk menilai persepsi siswa tentang iklim sekolah dan diberikan kepada siswa kemungkinan besar berlawanan dengan variabilitas antara sekolah. Mengabaikan variabilitas ‘antar-sekolah’ pada sejumlah siswa di sekolah-sekolah akan menghasilkan bias terhadap prediksi dalam parameter model analisis faktor. Oleh karena itu, diperlukan untuk memperpanjang metodologi *multilevel* dengan kerangka analisis faktor. (Kaplan, Kim & Kim, 2009). Untuk memulai, kita asumsikan bahwa vektor dari respon siswa dapat dinyatakan dalam *multilevel linear factor model* sebagai berikut:

$$y_{ig} = v + \Lambda_w \eta_{w_{ig}} + \Lambda_b \eta_{b_g} + \epsilon_{w_{ig}} + \epsilon_{b_g} \dots (7)$$

dimana y_{ig} sama seperti yang didefinisikan sebelumnya, v adalah *grand mean*, Λ_w adalah matriks muatan faktor untuk kelompok *within*, $\eta_{w_{ig}}$ merupakan faktor yang bervariasi secara acak di seluruh unit kelompok *within*, Λ_b adalah matriks muatan faktor kelompok *between*, η_{b_g} merupakan faktor yang bervariasi acak di seluruh kelompok, $\epsilon_{w_{ig}}$ dan ϵ_{b_g} merupakan *within* dan *between group uniqueness*. Menggunakan asumsi standar analisis faktor linear, di sini diperluas untuk kasus *multilevel*, matriks kovarians total didefinisikan dalam Persamaan (2) dapat dinyatakan dalam hal model faktor parameter sebagai:

$$\Sigma_T = \Lambda_w \Phi_w \Lambda_w' + \Theta_w + \Lambda_b \Phi_b \Lambda_b' + \Theta_b \dots (8)$$

dimana Φ_w dan Φ_b adalah matriks kovarians faktor untuk *within group* dan *between group* dan Θ_w dan Θ_b adalah matriks diagonal dari *unique variances* untuk bagian *within group* dan *between group*. Secara umum, biasanya mudah untuk menentukan struktur faktor untuk variabel *within school*. Hal ini juga mudah untuk memungkinkan variabel *within school* bervariasi *between school*. Kesulitan konseptual sering timbul untuk menjamin struktur faktor untuk menjelaskan variasi *between groups* (Kaplan, Kim & Kim, 2009). Sedangkan penelitian terbaru dari Geldhof, Preacher dan Zyphur (2014) menjelaskan bahwa MCFA merupakan ekstensi dari CFA yang dapat mengakomodir data *two-level* dan memungkinkan estimasi secara terpisah dan analisis matriks kovarians *within-* dan *between-cluster*.

Ada beberapa pendekatan dalam literatur metodologi untuk melakukan MCFA (mis, Muthén, 1994). terdapat metode baru yang dikembangkan oleh Muthén dan Asparouhov (2009, 2011) untuk melakukan analisis MSEM. MCFA adalah kasus khusus dari MSEM tanpa jalur struktural yang menghubungkan variabel laten, dalam banyak cara yang sama bahwa *single-level* CFA adalah kasus khusus dari SEM. Yuan dan Bentler (2007) menjelaskan berbagai keuntungan dari penggunaan metode *multilevel* SEM yaitu: (1) lebih mudah untuk mengetahui di *level* mana terjadi kesalahan spesifikasi; (2) indeks fit dari SEM konvensional

dapat dengan mudah dikembangkan untuk mengevaluasi model pada *level* yang terpisah dari model *multilevel*; (3) diagnosa pada model SEM konvensional dari berbagai literatur akan dapat diaplikasikan dengan mudah untuk mengecek kesalahan spesifikasi dari model *multilevel*; (4) kesalahan spesifikasi pada salah satu *level* tidak mempengaruhi secara sistematis evaluasi dari model pada *level* lainnya.

Secara singkat, model *Multilevel Confirmatory Factor Analysis* (MCFA) dijelaskan sebagai kasus khusus dari model SEM menurut Muthén dan Asparouhov (2009) yang model ini dijelaskan oleh satu set dari tiga persamaan (notasi mereka):

$$Y_{ik} = \Lambda_k \eta_{ik} \dots (7)$$

$$\eta_{ik} = \alpha_k + B_k \eta_{ik} + \zeta_{ik} \dots (8)$$

$$\eta_k = \mu + \beta \eta_{ik} + \zeta_k \dots (9)$$

di mana i dan kasus indeks k (unit *level 1*) dan *cluster* (unit *level 2*), masing-masing. Y_{ik} adalah vektor dari p variabel yang diukur; $\Lambda_k = \Lambda = [I_p \ 0_{p \times m} \ I_p \ 0_{p \times m}]$ adalah $(p \times (2p + 2m))$ *factor loading matrix* yang menghubungkan Y_{ik} ke bagian laten p di kedua tingkat *within-* dan *between-cluster* dan m faktor umum pada kedua tingkat; η_{ik} adalah vektor panjang $(2p + 2m)$ mengandung p laten bagian *within-cluster*, m yaitu faktor umum *within-cluster*, p laten bagian *between-cluster*, dan m faktor umum *between-cluster*; α_k adalah vektor panjang $(2p + 2m)$ yang berisi item *intercepts* p dan m faktor umum *between-cluster*; B_k adalah $(2p + 2m) \times (2p + 2m)$ matriks yang mengandung muatan faktor *within-cluster*; η_k ($r \times 1$) terdiri dari koefisien acak subskrip k dari α_k dan B_k , termasuk faktor umum *between-cluster*; μ ($r \times 1$) berisi *mean* dari koefisien tersebut dan item *intercepts* (jika diinginkan); β ($r \times r$) berisi muatan faktor *between-cluster*; ζ_{ik} berisi *unique factors* dan faktor umum residu untuk model *within-cluster*; dan ζ_k ($r \times 1$) mengandung *unique factors* dan faktor umum residu untuk model *between-cluster*. Akhirnya, $\zeta_{ik} \sim \text{MVN}(0, \psi_w)$, dan $\zeta_k \sim \text{MVN}(0, \psi_B)$.

Geldhof, Preacher & Zyphur (2014) mengemukakan bahwa, meskipun model dasar MCFA dapat dijabarkan dalam berbagai

cara, pada penelitian ini akan dibatasi hanya pada model faktor tanpa kovariat, hanya item kontinu, dan tidak ada *latent regression*. Selain itu, dalam kondisi ini hanya dipertimbangkan kasus di mana *item intercepts* dihilangkan, muatan faktor tidak berbeda secara acak di tingkat *cluster* ($B_k = B$), dan struktur faktor *configural* identik di seluruh tingkatan. Penyederhanaan ini menghasilkan bentuk Persamaan 7 dan 8 yang dibatasi menjadi:

$$Y_{ik} = \Lambda \eta_{ik} \dots (10)$$

$$\eta_{ik} = \alpha_k + B \eta_{ik} + \zeta_{ik} \dots (11)$$

Intraclass Correlation Coefficients (ICC)

Terdapat dua jenis dari ICC yang dapat dihitung: *latent factor* ICC dan *observed variable* ICC. (Hsu et al, 2016). *Latent factor* ICC merupakan hal yang lebih umum digunakan pada model *multilevel* CFA (Heck & Thomas, 2015; Muthen, 1994). dimana *latent factor* ICC dikomputasi dengan rumus:

$$\text{Latent Factor ICC} = \frac{B}{B + W} \dots (12)$$

Dimana B merupakan proporsi varians *latent factor* pada *between-level* dan W pada *within-level*. Namun perlu dicatat bahwa *latent factor* ICC hanya bisa dikomputasi saat struktur dari model sama pada *between* dan *within-level* (misal, asumsi struktur model identik) sehingga dapat dilakukan manipulasi terhadap besarnya *latent factor* ICC dengan cara menetapkan besaran varians faktor pada *within level* sebagai konstan dan mengubah-ubah besarnya varians faktor pada *between level*. (Hsu et al., 2016). Sedangkan ICC jenis kedua yaitu *observed variable* ICC dapat dikomputasi dengan rumus:

$$\text{Observed Variable ICC} = \frac{b}{b + w} \dots (13)$$

Dimana $b = (\text{muatan faktor } \textit{between-level})^2 \times \text{varians faktor } \textit{between-level} + \text{varians residual } \textit{between-level}$ dan $w = (\text{muatan faktor } \textit{within-level})^2 \times \text{varians faktor } \textit{within-level} + \text{varians residual } \textit{within-level}$. (Hsu et al., 2016).

Bias Estimasi

Menurut kamus *American Psychological Association* (2015) bias estimasi terjadi apabila-

la nilai yang diperoleh dari data sampel yang secara konsisten *underestimate* atau *overestimate* dari nilai sebenarnya dalam populasi berjumlah besar yang diteliti. Dengan kata lain, sebuah estimator dikatakan bias ketika rata-rata nilainya berbeda dari nilai parameter yang dimaksudkan untuk mewakili nilai yang sebenarnya. Hal ini disebut juga sebagai bias statistik. Marsh, Hau & Greyson (2005) menjelaskan bila nilai yang diharapkan dari statistik yang dilakukan pada sampel bervariasi secara sistematis dengan jumlah N, maka statistik yang dihasilkan akan menghasilkan bias dari parameter populasi sebenarnya.

Hal ini sejalan dengan penelitian sebelumnya (misalnya, Muthén, Kaplan & Hollis, 1987), yang juga mempertimbangkan ukuran mutlak presentase bias < 10% untuk mewakili tingkat bias yang dapat diterima. Karena estimasi parameter pada penelitian ini tidak diantisipasi akan terdistribusi secara normal, estimasi parameter median di setiap kondisi yang lebih baik dijadikan ukuran pemusatan statistik daripada *mean* estimasi parameter. Oleh karena itu bias dalam setiap kondisi dihitung sebagai $(\text{estimasi median} - \text{parameter}) / \text{parameter} \times 100$. (Geldhof, Preacher & Zyphur, 2014). Dalam penelitian lainnya, Cai (2010b) menggunakan dua ukuran keakuratan dari estimasi parameter yaitu bias estimasi dan *root mean squared error* (RMSE), dimana pada sebuah parameter θ , bias estimasi didefinisikan sebagai $M^{-1} \sum_{i=1}^M (\theta - \hat{\theta}_i)$, dimana M merupakan jumlah replikasi *Monte Carlo* keseluruhan dan $\hat{\theta}_i$ adalah MLE dari θ pada replikasi ke i. Sedangkan RMSE dirumuskan menjadi $\sqrt{M^{-1} \sum_{i=1}^M (\theta - \hat{\theta}_i)^2}$. Tetapi, dalam kondisi dimana terdapat bias sistematis yang besar, RMSE menjadi tidak informatif sebagai ukuran dari efisiensi karena akan didominasi oleh bias. (Preacher, Zhang & Zyphur, 2011).

Metode Penelitian

Penelitian ini menggunakan metode *Markov Chain Monte Carlo* (MCMC) yang berdasar pada pendekatan *Bayesian*. Metode

Monte Carlo merupakan penelitian dengan teknik simulasi di mana sampel dengan jumlah besar dan sifat tertentu yang ditentukan (misalnya, normalitas, ukuran, jenis model) dihasilkan oleh komputer untuk menilai penggunaan prosedur statistik atau parameter dalam berbagai kondisi. Sebagai contoh, seorang peneliti mungkin melakukan penelitian dengan metode *Monte Carlo* dengan sampel berjumlah besar yang terdistribusi normal dari berbagai ukuran sampel (misalnya, N = 50, 100, 200, 400, 800) di mana model struktural diterapkan untuk menggambarkan data. Hasil penelitian akan membantu peneliti menentukan kondisi di mana model berfungsi dengan benar (yaitu, sesuai dengan data) serta menunjukkan batas-batasnya (misalnya, tidak cocok dengan baik dengan sampel ukuran kurang dari 200). (*American Psychological Association*, 2015). Selanjutnya akan dijelaskan mengenai model populasi yang digunakan sebagai acuan proses pembangkitan data dan juga desain simulasi yang ditetapkan. Hal ini sebenarnya merupakan konsep-konsep terkait dasar-dasar pendekatan Bayesian seperti distribusi *prior*, *posterior* dan kaitannya dengan fungsi *likelihood*, namun penjelasannya tidak dijabarkan dengan rinci untuk memudahkan pelaporan. Penjelasan lengkap dapat dilihat pada berbagai literatur yang tersedia (misal, van De Schoot et al., 2014).

Model populasi yang digunakan pada penelitian ini merupakan model yang diambil dari penelitian Muthén (1994) yang berisi delapan indikator yang masing-masing diwakili satu faktor yaitu η_B pada *between level* dan η_W pada *within level*. Pada penelitian ini model MCFA pada bagian *between-level* memiliki struktur faktor yang muatannya berbeda dengan model *within-level*. Hal ini dilakukan karena mengikuti informasi yang didapat dari penelitian sebelumnya berupa data dan memilih untuk menggunakannya dibandingkan menciptakan kondisi baru dimana dilakukan variasi pada muatan faktor dan varians *residual* untuk menjadi model populasi sebagai proses pembangkitan data. Namun, berdasarkan model di atas, dapat disimpulkan bahwa model ini memiliki struktur

tur faktor yang identik baik pada *between-level* maupun *within-level*, sehingga *latent factor* ICC dapat digunakan dalam penelitian ini karena struktur faktor yang identik tersebut. Berdasarkan model ini, spesifikasi parameter bebas yang akan diestimasi dibagi menjadi dua bagian yaitu *between-level* dan *within level*. Spesifikasi parameter bebas pada *within-level* sebanyak 8 λ , 8 θ . Sedangkan spesifikasi parameter bebas pada *between-level* yaitu: 8 ν , 8 λ dan 8 θ . Sehingga keseluruhan berjumlah 40 parameter bebas yang akan diestimasi Adapun parameter bebas pada *between-level* dan *within-level* untuk melakukan simulasi data ditetapkan terlebih dahulu dimana hal utama yang ditetapkan adalah muatan faktor maupun varians *residual* yang digunakan sebagai *starting values* bagi estimasi parameter bebas yang dipaparkan pada Tabel 1.

Tabel 1. Muatan Faktor Model Populasi

Item	Muatan Faktor		Varians Residual	
	<i>Between</i>	<i>Within</i>	<i>Between</i>	<i>Within</i>
Y ₁	0,97	0,52	0,05	0,72
Y ₂	0,98	0,49	0,03	0,75
Y ₃	0,92	0,32	0,15	0,89
Y ₄	0,88	0,25	0,22	0,93
Y ₅	0,89	0,34	0,20	0,88
Y ₆	0,84	0,23	0,29	0,94
Y ₇	0,80	0,26	0,36	0,93
Y ₈	0,77	0,31	0,40	0,90

Berdasarkan informasi pada Tabel 1, untuk menciptakan kondisi ICC yang berbeda-beda, varians pada faktor *between-level* ditetapkan nilainya menjadi antara 0.06 sampai 0.26 sehingga menghasilkan kondisi ICC yang diinginkan yaitu 0.05, 0.10, 0.15, 0.20 dan 0.25 sebagai model awal yang digunakan untuk proses pembangkitan data untuk studi dengan simulasi data. Model ini merupakan model yang menjadi distribusi yang akan diuji pada berbagai kondisi yang diciptakan, adapun kondisi yang dimaksud, akan dijelaskan pada bagian selanjutnya.

Karena penelitian ini menggunakan metode simulasi data, maka harus ditetapkan terlebih dahulu kondisi simulasi data untuk melihat hasil dari simulasi berdasarkan kondisi yang berbeda-beda, kondisi yang

ditetapkan harus ditetapkan berdasarkan cara yang tersedia maupun dengan kriteria tertentu untuk dapat diaplikasikan, meskipun hal ini sebenarnya menjadi kritik dari aliran klasik. Adapun kondisi yang ditetapkan sebagai desain simulasi pada penelitian ini yaitu:

1. Jumlah *Cluster*
30, 50, 100 dan 150
2. Jumlah Observasi Pada *Cluster*
10, 20 dan 50
3. *Intraclass Correlation Coefficients* (ICC)
0.05, 0.10, 0.15, 0.20 dan 0.25
4. Metode Estimasi
Estimator yang digunakan pada penelitian ini adalah ML, MLF, MLR, WLSMV dan BAYES.

Dengan kondisi simulasi yang ditetapkan $4 \times 3 \times 5 \times 5$ akan menghasilkan kondisi sebanyak 300 yang masing-masing akan direplikasi sebanyak 1000 kali. Penentuan kondisi ini juga diikuti dengan observasi maupun komparasi dengan keadaan pada kondisi nyata dimana kondisi-kondisi tersebut umum digunakan pada aplikasinya dengan data yang didapat dari lapangan. Hal ini dilakukan agar hasil yang didapat dari penelitian ini yang dilakukan menggunakan studi dengan simulasi dapat diperbandingkan ataupun dikaitkan dengan pengaplikasiannya yang telah umum dilakukan.

Hasil Penelitian

Akan dipaparkan pengelompokan hasil estimasi yang menunjukkan besaran bias pada masing-masing kondisi ICC yaitu 0.05, 0.10, 0.15, 0.20 dan 0.25 yang membandingkan lima metode estimasi yaitu ML (*maximum likelihood*), MLF (*full information maximum likelihood*), MLR (*maximum likelihood robust*), WLSMV (*robust weighted least square*) dan BAYES (*Bayesian estimation*). Adapun pengelompokannya dijelaskan masing-masing pada grafik yang akan ditampilkan.

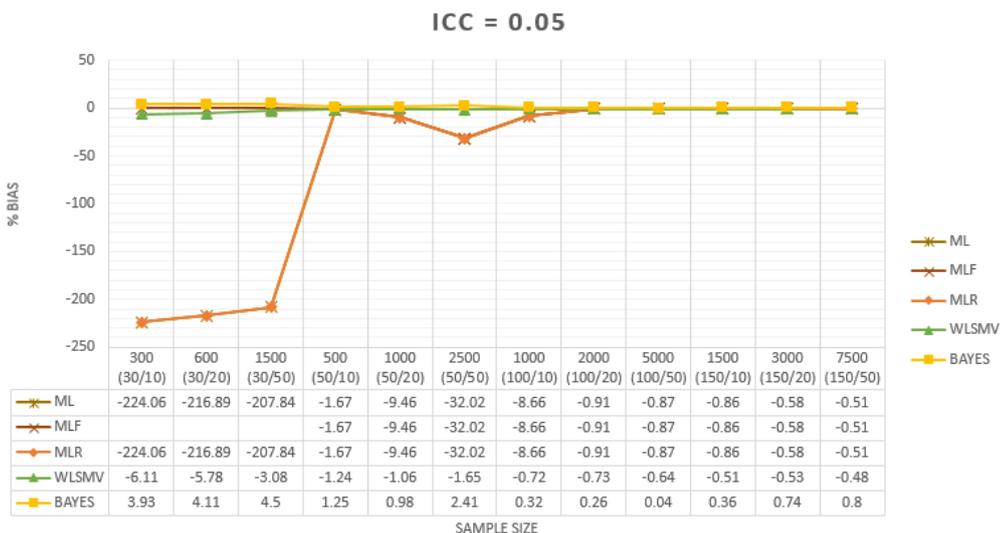
Pengelompokan hasil simulasi data berdasarkan kondisi dengan ICC sebesar 0.05 pada kondisi ukuran sampel maupun kelima metode estimasi yang digunakan yaitu ML, MLF, MLR, WLSMV dan BAYES digambarkan dalam grafik yang masing-masing berisikan 60 kondisi yang memuat in-

formasi mengenai besaran bias yang terjadi pada kondisi ICC ini, adapun grafik yang dimaksud berada pada Gambar 1.

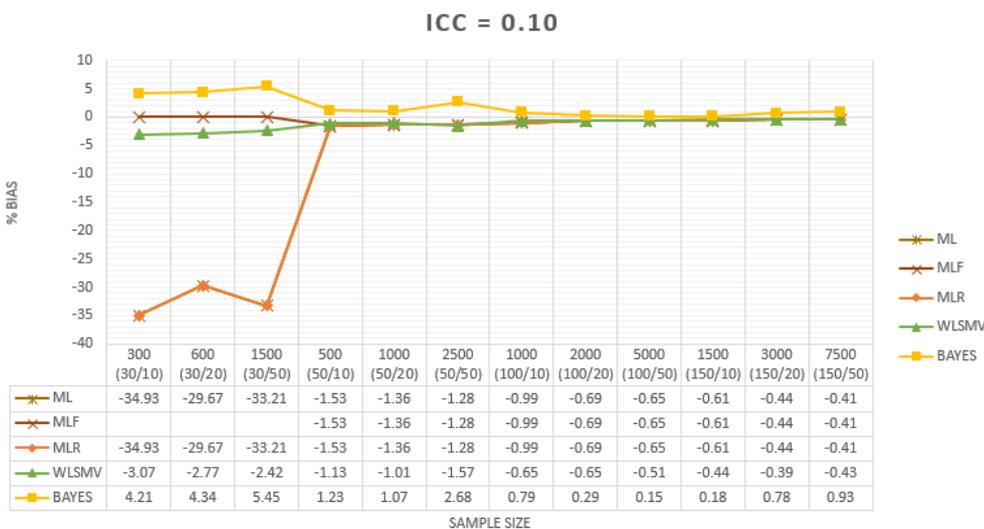
Berdasarkan grafik pada Gambar 1, dengan kondisi ICC sebesar 0.05 terlihat bahwa metode estimasi ML dan MLR memberikan hasil estimasi yang bias pada kondisi sampel 30/10, 30/20, 30/50 dan 50/50.

Sedangkan pada kondisi sampel lainnya tidak terjadi bias yang melebihi batas yang telah ditetapkan. Lalu, dengan metode estimasi BAYES dan WLSMV hasil estimasi parameter yang diberikan berbeda dengan ML & MLR pada kondisi sampel yang sama dimana metode estimasi BAYES dan WLS-

MV memberikan hasil estimasi yang lebih baik dan tidak bias pada kondisi ukuran sampel manapun. Sedangkan metode estimasi MLF tidak berfungsi pada kondisi sampel 30/10, 30/20 dan 30/50 dimana meskipun metode estimasi ini sama-sama berbasis *maximum likelihood* seperti ML dan MLR tetapi proses kalkulasi yang berbeda menganggap bahwa kondisi ukuran sampel tersebut tidak memenuhi syarat untuk melakukan estimasi parameter. Namun pada kondisi sampel yang lain, metode estimasi MLF dapat digunakan dan memberikan hasil yang sama dengan ML ataupun MLR.



Gambar 1. Bias pada Kondisi dengan ICC 0.05



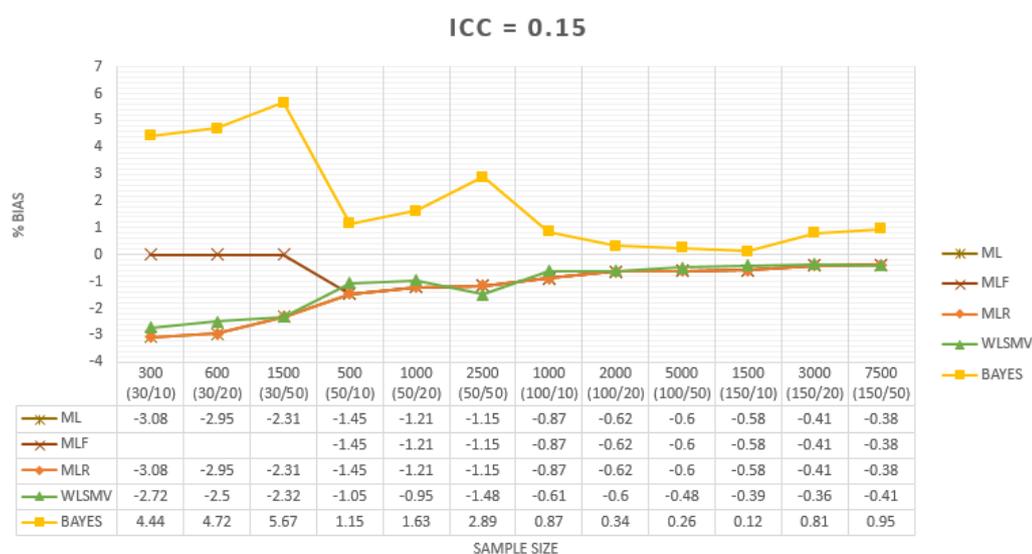
Gambar 2. Bias pada Kondisi dengan ICC 0.10

Berdasarkan grafik di Gambar 2, dengan kondisi ICC sebesar 0.10 terlihat bahwa metode estimasi ML dan MLR memberikan hasil estimasi yang bias pada kondisi sampel 30/10, 30/20 dan 30/50. Sedangkan pada kondisi sampel lainnya tidak terjadi bias yang melebihi batas yang telah ditetapkan. Lalu, dengan metode estimasi BAYES dan WLSMV hasil estimasi yang diberikan berbeda dengan ML & MLR pada kondisi sampel yang sama dimana metode estimasi BAYES dan WLSMV memberikan hasil estimasi yang lebih baik dan tidak bias melebihi kriteria yang ditetapkan pada kondisi ukuran sampel manapun. Sedangkan metode estimasi MLF tidak berfungsi pada kondisi sampel 30/10, 30/20 dan 30/50 dimana meskipun metode estimasi ini sama-sama berbasis *maximum likelihood* seperti ML dan MLR tetapi proses kalkulasi yang berbeda menganggap bahwa kondisi dengan ukuran sampel tersebut tidak memenuhi syarat untuk melakukan estimasi parameter. Namun pada kondisi sampel yang lain, metode estimasi MLF dapat digunakan dan memberikan hasil yang sama dengan ML ataupun MLR.

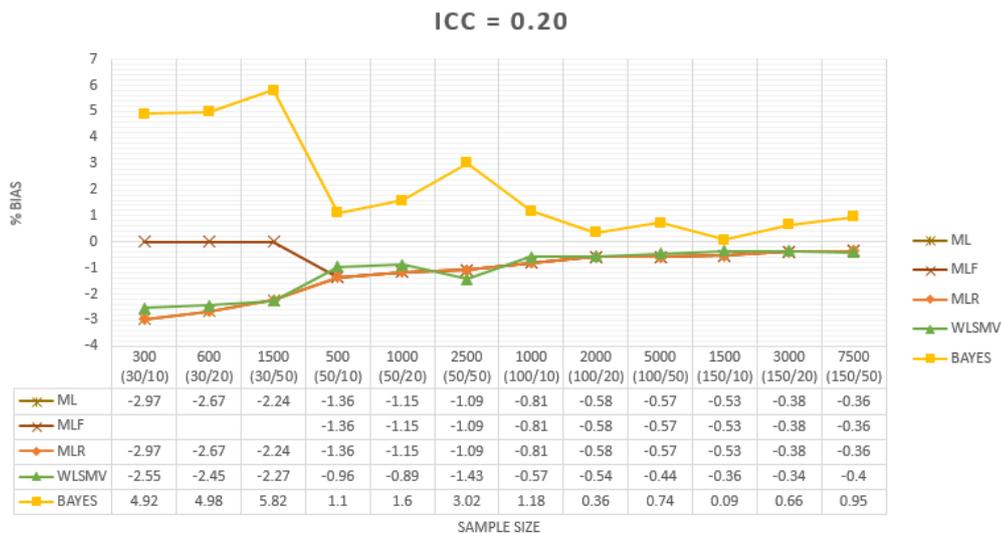
Berdasarkan grafik pada Gambar 3, secara keseluruhan pada kondisi ICC 0.15 pada seluruh kondisi ukuran sampel tidak terjadi bias yang melebihi batas yang dite-

apkan yaitu $> 10\%$. Berdasarkan arah dari bias meskipun sangat kecil, pada metode estimasi ML, MLR, WLSMV bias yang terjadi arahnya negatif dan pada metode estimasi BAYES arahnya positif dan lebih besar dibandingkan metode yang berbasis *maximum likelihood*. Tetapi terdapat pengecualian dimana MLF tidak dapat berfungsi dengan baik pada kondisi sampel yang tidak memenuhi syarat. Terdapat kesamaan pola dimana semakin besar jumlah sampel maka semakin rendah bias yang terjadi seperti dapat dilihat pada gambar 3.

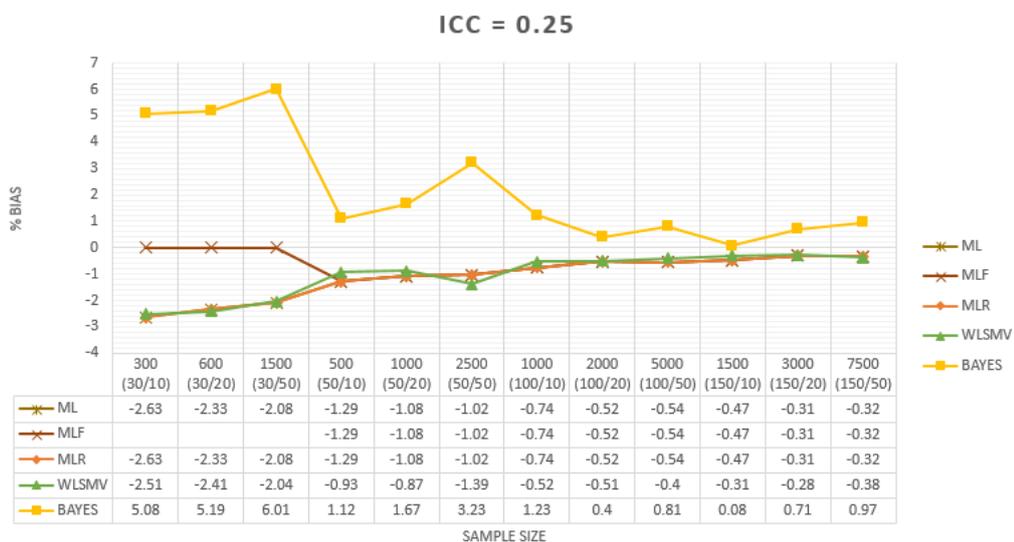
Berdasarkan grafik di Gambar 4, secara keseluruhan pada kondisi ICC 0.20 pada seluruh kondisi ukuran sampel tidak terjadi bias yang melebihi batas yang ditetapkan yaitu $> 10\%$. Berdasarkan arah dari bias meskipun sangat kecil, pada metode estimasi ML, MLR, WLSMV bias yang terjadi arahnya negatif dan pada metode estimasi BAYES arahnya positif dan lebih besar dibandingkan metode yang berbasis *maximum likelihood*. Tetapi terdapat pengecualian dimana MLF tidak dapat berfungsi dengan baik pada kondisi sampel yang tidak memenuhi syarat. Terdapat kesamaan pola dimana semakin besar jumlah sampel maka semakin rendah bias yang terjadi seperti dapat dilihat pada grafik sebelumnya.



Gambar 3. Bias pada Kondisi Dengan ICC 0.15



Gambar 4. Bias pada Kondisi dengan ICC 0.20



Gambar 5. Bias pada Kondisi dengan ICC 0.25

Berdasarkan grafik di Gambar 5, secara keseluruhan pada kondisi ICC 0.25 pada seluruh kondisi ukuran sampel tidak terjadi bias yang melebihi batas yang ditetapkan yaitu $> 10\%$. Berdasarkan arah dari bias meskipun sangat kecil, pada metode estimasi ML, MLR, WLSMV bias yang terjadi arahnya negatif dan pada metode estimasi BAYES arahnya positif dan lebih besar dibandingkan metode yang berbasis *maximum likelihood*. Terdapat kesamaan pola dimana semakin besar ukuran sampel maka semakin rendah bias yang terjadi seperti dapat dilihat pada grafik sebelumnya.

Hal ini menunjukkan bahwa ke lima metode estimasi yang digunakan merupakan metode estimasi yang tidak bias, konsisten dan efisien untuk mengestimasi parameter data yang berstruktur *multilevel* apabila ukuran sampel dari data memenuhi kriteria yang telah ditetapkan oleh penelitian-penelitian terdahulu terkait ukuran sampel yang dibutuhkan meskipun terdapat pengecualian dimana pada kondisi sampel tertentu MLF tidak dapat berfungsi dengan baik. Berlawanan dengan metode estimasi MLR, metode estimasi BAYES merupakan metode yang memiliki kelebihan untuk melakukan estimasi

asi pada kondisi data yang tidak ideal yaitu berupa ukuran sampel kecil ataupun ICC yang rendah. Metode estimasi BAYES juga membutuhkan waktu komputasi yang lebih panjang dibanding metode estimasi ML, MLF, MLR & WLSMV. Kelima metode ini merupakan metode yang memiliki dasar yang berbeda satu sama lain sehingga memiliki kelebihan dan kekurangan masing-masing, tetapi secara keseluruhan kelima metode ini merupakan metode estimasi yang baik untuk digunakan dalam mengestimasi parameter data berstruktur *multilevel*. Selanjutnya, untuk melakukan perbandingan hasil yang didapat dari penelitian ini dengan penelitian sebelumnya yang relevan, akan dijelaskan lebih lanjut pada bagian selanjutnya pada subbab diskusi.

Pembahasan

Berdasarkan hasil analisis data yang dilakukan, penelitian ini menunjukkan bahwa terdapat berbagai temuan yang menarik pada analisis data yang berstruktur hirarki. Analisis data *multilevel* adalah topik yang kompleks karena mengacu pada kontribusi dari berbagai bidang metodologi penelitian. Dua perspektif dapat membedakan, yaitu pengambilan sampel dan parameter yang berbeda-beda. Dari perspektif *sampling*, data yang bertingkat dapat dilihat seperti yang diperoleh dengan *cluster sampling*. Analisis perlu untuk menentukan variasi stokastik yang mencerminkan skema *sampling*, seperti membuat model yang terurai dari variasi kelompok (*cluster*) dan komponen individu. (Muthén, 1994).

Hasil pengambilan sampel menggunakan metode *multistage sampling* pada data hirarki yang terstruktur (misalnya, siswa berkumpul di dalam satu unit kelas), membuat residual bergantung pada variasi *between-cluster*. Skor pada variabel yang diamati dari anak-anak dalam satu kelas mungkin lebih mirip daripada anak-anak di kelas yang berbeda, misalnya. Apabila mengabaikan struktur data hirarkis dapat terjadi bias pada estimasi hubungan antar item. (Geldhof, Preacher & Zyphur, 2014), ICC bersamaan dengan ukuran sampel dan juga informasi

jumlah individu dalam *cluster* merupakan informasi yang penting dalam permodelan *multilevel* (Stapleton, Yang & Hancock, 2016). Penelitian tersebut menunjukkan bahwa ICC merupakan faktor penting yang perlu diperhatikan dalam menganalisis data *multilevel*, dimana hasil dari penelitian ini menunjukkan bahwa kondisi ICC berpengaruh terhadap bias estimasi parameter data *multilevel*.

Terdapat temuan yang menarik pada kondisi ICC 0.05 dimana apabila dilakukan perbandingan pada dua kondisi sampel yaitu 50/20 dan 100/10 yang sama-sama sebesar 1000 observasi terjadi bias pada kondisi 50/20 sebesar -9.46 tetapi biasanya menurun pada kondisi 100/10 yaitu sebesar -8.66, hal ini menunjukkan bahwa struktur dari sampel mempengaruhi bias dimana jumlah *cluster* yang lebih besar membuat estimasi parameter data *multilevel* menjadi lebih baik dibanding *cluster* yang lebih sedikit. Hal ini sejalan dengan pendapat Cronbach (1976, dalam Muthén, 1991) yaitu jika terbentur permasalahan, mungkin lebih baik mengamati siswa per kelas yang lebih sedikit untuk membantu menambah kelas yang lebih banyak dimana kelas sebagai *cluster* yang jumlahnya lebih banyak akan memberikan hasil estimasi yang terbukti lebih baik dan tidak bias. Besaran bias yang terjadi pada kondisi dengan sampel kecil seperti 30/10, 30/20 & 30/50 pada metode estimasi MLR menunjukkan bahwa MLR tidak berfungsi dengan baik apabila ICC kecil dan juga jumlah individu dalam cluster lebih kecil dari 50 sesuai dengan hasil penelitian terdahulu. (Hox & Mass, 2004, dalam Stapleton, Yang & Hancock, 2016).

Hasil penelitian ini juga sejalan dengan penelitian Pornprasertmanit, Lee & Preacher (2014) yang menyatakan bahwa bias akan semakin rendah saat jumlah *cluster* semakin besar apabila menggunakan metode estimasi MLR. Temuan ini juga sejalan dengan pendapat Heck & Thomas (2015) yang menyatakan bahwa ukuran sampel pada tingkat grup (jumlah grup) umumnya lebih penting dari ukuran sampel tingkat individu (jumlah observasi dalam *cluster*) yang sejalan dengan Łaszkiewicz (2013) dan

Cronbach (1976, dalam Muthen, 1991) yang berpendapat bahwa hal yang penting bukan jumlah yang besar pada pengamatan per unit namun jumlah besar kelompok tampaknya lebih penting untuk mendapatkan estimasi yang akurat.

Temuan ini juga membuktikan kesimpulan yang diambil dari penelitian sebelumnya oleh Preacher, Zhang & Zyphur (2011) yaitu semakin meningkatnya jumlah grup, ukuran *cluster* dan ICC yang berada dalam batas menunjukkan hasil yang baik pada penggunaan metode MSEM terkhusus pada bias yang terjadi dengan metode estimasi MLR. Hal ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa jika jumlah *cluster* besar dan ICC bernilai setidaknya 0.10, MSEM merupakan metode yang paling efisien seiring bertambahnya jumlah *cluster* dan metode ini memberikan estimasi yang lebih efisien dibandingkan penggunaan nilai rata-rata dari kelompok apabila ukuran kelompok yang besar digunakan (Lüdtke, Marsh, Robitzsch, Trautwein, Asparouhov, & Muthén, 2008).

Pada kondisi dengan ukuran sampel kecil sebesar 30/10, 30/20, 30/50 dan 50/50 metode estimasi MLR memberikan hasil estimasi yang bias, hal ini sejalan dengan penelitian terdahulu yang menyatakan bahwa dengan ukuran sampel yang kecil, metode ML menghasilkan estimasi yang bias (Morris, 1995 dalam Heck & Thomas, 2015). Lebih lanjut, untuk mengatasi permasalahan ukuran sampel yang kecil, metode estimasi *Bayesian* merupakan pendekatan alternatif yang dapat digunakan (Heck & Thomas, 2015) yang pada penelitian ini terbukti bahwa dalam kondisi ukuran sampel kecil, metode estimasi *Bayesian* memberikan hasil estimasi yang lebih baik dari MLR dengan ukuran bias yang masih dapat diterima. Temuan ini juga membuktikan hasil penelitian sebelumnya yang menyatakan bahwa pada kondisi data yang tidak ideal, metode estimasi BAYES dapat bekerja dengan baik (Lindley & Smith, 1972; Smith, 1973; Morris, 1995; dalam Heck & Thomas, 2015). Metode estimasi BAYES juga memberikan keuntungan dimana model *multilevel* SEM

sulit diestimasi dengan metode estimasi berbasis ML karena membutuhkan jumlah integrasi yang banyak, sehingga metode estimasi BAYES sangat baik untuk digunakan (Asparouhov & Muthen, 2014).

Pada kondisi ukuran sampel yang lain selain ukuran sampel 30/10, 30/20, 30/50 dan 50/50 metode estimasi ML, MLF, MLR terbukti konsisten, *unbiased* dan efisien dimana hasil penelitian ini menunjukkan bahwa semakin besar ukuran *cluster* dalam kondisi ICC lebih besar dari 0.05, maka semakin kecil bias yang terjadi pada estimasi parameter data *multilevel*. Sedangkan metode estimasi BAYES memberikan hasil yang berbeda, dimana metode ini membutuhkan waktu komputasi yang lebih lama dibandingkan penggunaan metode estimasi MLR.

Penelitian ini juga menjawab permasalahan yang terkait dengan penggunaan metode estimasi MLF dimana seharusnya metode berbasis *maximum likelihood* seperti ML, MLF dan MLR menghasilkan hasil yang sama, tetapi pada penelitian ini terdapat temuan bahwa pada beberapa kondisi ukuran sampel dan ICC tertentu MLF tidak dapat berfungsi. Hal ini terjadi karena menurut penelitian terdahulu, bagaimanapun untuk ukuran sampel yang kecil/menengah MLF akan mengalami *overestimate* pada *standard error*-nya. Hal ini biasanya terjadi pada kasus dimana perbandingan antara jumlah unit independen (observasi pada model satu *level* atau *cluster* pada model *multilevel*) dan jumlah parameter kurang dari 10. (Asparouhov & Muthen, 2012). Dalam penelitian ini terdapat gambaran jumlah ukuran sampel yang pasti dimana MLF tidak dapat berfungsi yaitu ukuran sampel 30/10, 30/20 dan 30/50. Sehingga pada penelitian mendatang untuk melakukan analisis MLLVM tidak disarankan untuk menggunakan metode estimasi MLF pada ukuran sampel tersebut.

Terkait penggunaan metode estimasi WLSMV pada penelitian ini, hasil yang didapat menunjukkan jika performa dari WLSMV lebih baik dari metode estimasi berbasis *maximum likelihood* bila tidak terjadi bias yang melebihi kriteria yang ditetapkan pada seluruh kondisi ukuran sampel dan ICC yang

tersedia. Hal ini sejalan dengan penelitian yang menunjukkan bahwa WLSMV menghasilkan estimasi yang lebih baik dibandingkan estimator berbasis ML pada seluruh kondisi sampel (Beauducel & Herzberg, 2006).

Sejalan dengan pendapat yang dikemukakan Meuleman & Billiet (2009) studi ini menggambarkan penggunaan studi dengan simulasi *Monte Carlo* yang berperan baik dalam menggambarkan akurasi dari estimasi yang dilakukan. Tetapi pendekatan *Bayesian* yang banyak menawarkan keuntungan dalam penggunaannya pada ukuran data yang kecil perlu diperhatikan bahwa efisiensi dari estimasi bergantung pada informasi yang tersedia sebagai *prior* dari parameter, tanpa informasi ini, estimasi yang dilakukan tidak akan memberikan hasil yang lebih baik dibandingkan metode estimasi lain dan bahkan dapat menghasilkan hasil yang lebih buruk (Heck & Thomas, 2015).

Terdapat berbagai keterbatasan dari penggunaan metode MSEM seperti permasalahan tentang konvergensi dan waktu komputasi, tetapi metode estimasi tingkat *advance* dan kekuatan alat untuk melakukan komputasi akan menyelesaikan permasalahan ini karena kompleksitas dari model ataupun desain dari penelitian yang berkembang akan terus memaksa untuk melakukan pengembangan pada prkateknya dimana para peneliti akan bereksperimen mengenai model-model yang akan di estimasi tetapi tetap berpegang pada bagaimana penelitian tersebut mungkin dilakukan dalam prakteknya. (Preacher, Zhang & Zyphur, 2016). Studi dengan simulasi data menggunakan metode MCMC menjawab persoalan untuk bagaimana efisien dan komputasi berdasar pada algoritma dengan bantuan komputer. (Cai, 2010b). Tetapi, tidak dapat dipungkiri juga apabila waktu yang dibutuhkan dalam proses komputasi *Bayesian* meningkat karena teknik dengan *iterative sampling* digunakan. (Van de Schoot, Kaplan, Denissen, Asendorpf, Neyer & Aken, 2014).

Hasil penelitian ini sejalan dengan penelitian Muthen & Asparouhov (2012) yang memperkenalkan aplikasi pendekatan *Bayesian* dalam penggunaannya pada SEM yang ber-

nama BSEM yang merupakan metode yang mudah dan cepat untuk menganalisis keterkaitan antara *loading* dimana pengaplikasian metode estimasi *Bayes* pada penelitian ini menghasilkan estimasi yang baik pada kondisi ICC dan juga ukuran sampel manapun. Namun, terdapat kelemahan yang sama seperti penelitian tersebut dimana analisis yang melibatkan kovarians *residual* menyebabkan komputasi yang sangat sulit karena konvergensi dari MCMC akan menjadi lambat.

Penelitian ini memberikan contoh bagaimana hipotesis yang dibuat dari distribusi *prior* informatif yang diuji dengan pendekatan *Bayesian* sejalan dengan penelitian sebelumnya yang menjelaskan bahwa hipotesis pada pendekatan klasik dan hipotesis informatif merupakan hal yang berbeda. bukan hanya dari bagaimana hipotesis tersebut dirumuskan namun juga bagaimana hipotesis tersebut dievaluasi. Hipotesis nol pada pendekatan klasik dapat dievaluasi menggunakan *p-value* sedangkan hipotesis informatif dapat dievaluasi menggunakan seleksi model *Bayesian*, hasil utama dari seleksi model *Bayesian* adalah distribusi *posterior* (Hojtink, Klugkist & Boelen, 2008). Tetapi pendekatan ini dikritik karena mengasumsikan bahwa setiap parameter memiliki distribusi pada populasi, bahkan termasuk kovarians dimana aliran *frequentist* tidak setuju dengan asumsi ini karena mereka berasumsi bahwa pada populasi hanya ada satu nilai tetap dari parameter (Van de Schoot, Kaplan, Denissen, Asendorpf, Neyer & Aken, 2014).

Simpulan

Berdasarkan hasil penelitian dan pembahasan, dapat disampaikan simpulan sebagai berikut. Hasil penelitian ini menunjukkan bahwa bias yang terjadi dipengaruhi oleh ukuran sampel dan ICC, penelitian ini juga menunjukkan bahwa metode estimasi WLSMV dan BAYES berfungsi lebih baik pada berbagai kondisi dibandingkan dengan metode estimasi berbasis ML.

Daftar Pustaka

American Psychological Association.
(2015). *APA dictionary of psychology* (2nd

- ed.). Washington, DC: American Psychological Association
- Asparouhov, T., & Muthén, B. O. (2003). Full-information maximum-likelihood estimation of general two-level latent variable models with missing data. *Mplus Working Paper*. Los Angeles, CA: Muthén & Muthén, Inc
- Asparouhov, T., & Muthén, B. O. (2012). Saddle points. *Mplus Working Paper*. Los Angeles, CA: Muthén & Muthén, Inc
- Asparouhov, T., & Muthén, B. O. (2014). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. *Mplus Working Paper*. Los Angeles, CA: Muthén & Muthén, Inc
- Beauducel, A. & Herzberg, P. Y. (2006) On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA, *Structural Equation Modeling: A Multidisciplinary Journal*, 13(2), 186-203
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of research in education*, 8(1), 158-233
- Cai, L. (2010b). Metropolis-hastings robbins-monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35(3), 307-335
- de Leeuw, J. & Meijer, E. (2008), *Handbook of multilevel analysis*. New York, NY: Springer.
- Dyer, N. G., Hanges, P. J., Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*. 16, 149–167
- Feinberg, R. A. & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, 35(2), 36-49
- Geldhof, G. J., Preacher, K. J., Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72–91
- Goldstein, H. (2011). *Multilevel Statistical Models* (4th ed.). London: Wiley.
- Hayes, A. F. (2006). A primer on multilevel modeling. *Human communication research*, 32, 385–410
- Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques* (3rd ed.). New York, NY: Routledge.
- Hoijtink, H., Klugkist, I., & Boelen, P. A. (2008). *Bayesian evaluation of informative hypotheses*. New York, NY: Springer
- Hsu, H., Lin, J., Kwok, O., Acosta, S., & Willson, V. (2016). The impact of intraclass correlation on the effectiveness of level-specific fit indices in multilevel structural equation modeling: a monte carlo study. *Educational and Psychological Measurement*, 1-27
- Julian, M. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, 8, 325–352.
- Kaplan, D., Kim, J. S., and Kim, S. Y. (2009). Multilevel latent variable modeling: current research and recent developments. in r. e. millsap & a. maydeu-olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 592-613). Thousand Oaks, CA: Sage
- Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- Łaszkiwicz, E. (2013). Sample size and structure for multilevel modelling: monte carlo investigation for the balanced design. *Quantitative methods in*

- economics*, 14(2), 19-28
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13, 203-229
- Marsh, H. W., Hau, K-T & Grayson, D. (2005). Goodness of Fit Evaluation in Structural Equation Modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary Psychometrics. A Festschrift for Roderick P. McDonald* (pp. 275-340). Mahwah NJ: Erlbaum.
- Mass, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92
- Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM?. *Survey Research Methods*, 3, 45-58.
- Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators, *Psychometrika*, 49, 115-132.
- Muthén, B. O. (1989). Latent variable modeling in heterogenous populations, *Psychometrika*, 54, 557-585.
- Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28(4), 338-354
- Muthén, B. O. (1994). Multilevel covariance structure analysis, *Sociological Methods & Research*, 22, 376
- Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel data, *Sociological Methodology*, 27, 453-480
- Muthén, B. O., & Asparouhov, T. (2009). Growth mixture modeling: analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis*. New York, NY: Taylor & Francis.
- Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313-335
- Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52, 431-462.
- Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide: Statistical analysis with latent variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, B. O. & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, 25, 267-316.
- Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring clustering in confirmatory factor analysis: some consequences for model fit and standardized parameter estimates. *Multivariate Behavioral Research*, 49, 518-543
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel sem. *Structural Equation Modeling*, 18, 161-182
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods*, 21 (2), 189-205

- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Snijders, T. A. B. (2005). Power and Sample Size in Multilevel Linear Models. In: B. S. Everitt and D.C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science*. Volume 3, 1570–1573. Chicester, UK: Wiley.
- Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, 20 (10), 1-40
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to research in child development. *Child Development*, 85, 842-860.
- Yuan, K-H. & Bentler, P. M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models, *Sociological Methodology*, 31(1), 53-82