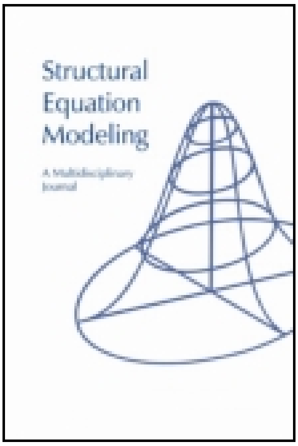

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches

Jiun-Yu Wu [a] & Oi-man Kwok [b]
[a] Institute of Education, National Chiao Tung University , Hsinchu , Taiwan
[b] Department of Educational Psychology , Texas A&M University
Published online: 23 Jan 2012.

Structural Equation Modeling, 19:16–35, 2012

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2012.634703

Psychology Press
Taylor & Francis Group

Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches

Jiun-Yu Wu
Institute of Education, National Chiao Tung University, Hsinchu, Taiwan

Oi-man Kwok
Department of Educational Psychology, Texas A&M University

Both ad-hoc robust sandwich standard error estimators (design-based approach) and multilevel analysis (model-based approach) are commonly used for analyzing complex survey data with nonindependent observations. Although these 2 approaches perform equally well on analyzing complex survey data with equal between- and within-level model structures (B. O. Muthén & Satorra, 1995), the performances of these 2 approaches for analyzing multilevel data with unequal between- and within-level structures have not yet been systematically examined. In this study, we extended B. O. Muthén and Satorra's (1995) study by comparing these 2 approaches and an additional model-based maximum model for analyzing multilevel data considering number of clusters, cluster size, intraclass correlation, and the equality of different level structures. The simulation results showed the model-based maximum model generally performed well across conditions. This model is also recommended as an alternative for analyzing nonindependent survey data, especially when the information of the higher level model structure is not known.

Keywords: complex survey data, design-based approach, maximum model, model-based approach, multilevel SEM

The cluster sampling or multistage sampling technique is widely used in educational research for its efficiency in time and resources. Unlike simple random sampling (SRS), which randomly selects a sample from a target population to ensure independence of observations, cluster sampling randomly samples naturally occurring groups and clusters of individuals or observations (Gall, Gall, & Borg, 2003; Stapleton, 2006). Data collected using cluster sampling

Correspondence should be addressed to Oi-man Kwok, Department of Educational Psychology, 4225 TAMU, College Station, TX 77843-4225, USA. E-mail: omkwok@neo.tamu.edu or Jiun-Yu Wu, Institute of Education, National Chiao Tung University, Hsinchu, Taiwan. E-mail: jiunyu.rms@gmail.com

tend to have correlated observations within clusters. For example, students from the same classroom are more likely to respond in a similar way because of the influence from the same environment. Conventional statistical methods that assume independent observations should not be used with data collected from cluster sampling due to the potential of nonindependent observations. The use of conventional statistical methods on nonindependent data can result in biased estimations of standard errors and incorrect statistical conclusions (Hox, 2002; Kalton, 1997; Kish, 1965).

In structural equation modeling, data are typically assumed to be collected through SRS and to be independently and identically distributed (du Toit & du Toit, 2008; Stapleton, 2006). However, in social and educational research, data with a hierarchical structure are not uncommon, especially when data are obtained through cluster sampling or multistage sampling. In these sampling techniques data are characterized by dependency among observations, such as students nested within schools or individuals nested within households (Lee, Forthofer, & Lorimor, 2006; Skinner, Holt, & Smith, 1989). "By ignoring the hierarchical structure of the data, incorrect parameter estimates, standard errors, and inappropriate fit statistics may be obtained" (du Toit & du Toit, 2008, p. 456).

Three analytic approaches are usually employed for analyzing data collected through cluster sampling, namely, disaggregated analysis, aggregated analysis, and multilevel modeling (Hofmann, 1997; Klein & Kozlowski, 2000). The first approach, disaggregated analysis, ignores higher level data structures (e.g., classroom level) and only models observations at the lower level structure (e.g., student level). This approach has been criticized for violating the assumption of independency under SRS (Hofmann, 1997; Raudenbush & Bryk, 2002). Neglecting the dependency among observations will generally result in underestimating fixed effects standard errors and lead to an inflated Type I error rate (De Leeuw & Kreft, 1995; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

Conversely, the second approach, aggregated analysis, as its name suggests, only analyzes aggregated data from the lower or individual level. Studies have shown that regression analysis performed on aggregated data can result in biased parameter estimates and underestimated standard errors associated with fixed effects (Croon & van Veldhoven, 2007; Lüdtke et al., 2008). Moreover, aggregated data cannot fully reflect the individual level variation (Au & Cheung, 2004; Klein et al., 2001). Studies have shown that neither approach—disaggregated analysis or aggregated analysis—can adequately reveal the complete picture of the relations between different levels of variables in multilevel data (Holt, Scott, & Ewings, 1980; Raudenbush & Bryk, 2002).

The third approach is to take the multilevel data sampling scheme into account, using either design-based or model-based approaches. The design-based approach takes the multilevel data or dependency into account by adjusting for parameter estimate standard errors based on the sampling design. The model-based approach analyzes the multilevel data by specifying a level-specific model for each data level. For example, for two-level clustered sampling data, the design-based approach analyzes the data with only one overall model and adjusts parameter estimate standard errors based on the sampling design, whereas the model-based approach analyzes the data by specifying (different) within-level and between-level models, respectively. Nevertheless, the design-based approach is commonly used by substantive researchers (e.g., Agrawal & Lynskey, 2007; Davidov et al., 2006; Hox & Kleiboer, 2007; Mathews et al., 2009; B. O. Muthén & Asparouhov, 2006). The design-based approach is seemingly to be often preferred given that it only requires specification of one single model, and researchers might be primarily interested in examining the within-level model.

Although the design-based approach is relatively simpler for model specification, it presumes that the within-level and between-level models are exactly the same. This assumption might not be always true. The advantage of the multilevel model is the flexibility for specifying different models with more than one data level. Indeed, B. O. Muthén and Satorra (1995) showed that these two approaches (design-based vs. model-based) performed equally well when analyzing complex survey data with exactly the same model structure for all data levels. However, the following design factors were not considered in their simulation: the structure or equality of the within- and between-level models, evaluation criteria (e.g., the coverage of the parameter estimates), and the empirical power for detecting the parameter estimates.

In this study, we extended B. O. Muthén and Satorra's (1995) findings by comparing the design-based and model-based approaches for analyzing multilevel data considering the following set of design factors: number of clusters, cluster size, intraclass correlation (ICC), and the structure/equality of the between-level and within-level models. We adopted M*plus* (version 6.0; L. K. Muthén & Muthén, 1998–2010) for all data generations and analyses.

M*plus* (version 6.0) has built-in routines (i.e., TYPE=COMPLEX and TYPE=TWOLEVEL) for analyzing multilevel data with the two approaches. The TYPE=COMPLEX routine is used for the design-based approach in which only one (single-level) model is needed for specification. The TYPE=TWOLEVEL routine is used for the model-based approach. The model-based approach allows researchers to specify models with different levels of data. By default, both routines use the maximum likelihood parameter estimator and the robust sandwich standard error estimator. The formula for calculating the variance components using the robust estimator includes a score factor "sandwiched" between two copies of the Hessian matrix (Hardin & Hilbe, 2007). In M*plus,* this estimation procedure is called the Maximum Likelihood Estimation with robust standard error correction (MLR). The MLR is useful for nonnormality and nonindependence of observations. The corresponding chi-square test statistic is asymptotically equivalent to the Yuan-Bentler T2* test statistic (L. K. Muthén & Muthén, 1998–2010). The robust parameter estimator, widely used in survey statistics, is also known as the Huber–White robust standard error estimate, survey variance estimate, design-based variance estimate, and empirical variance estimate (Hardin & Hilbe, 2007). Using this robust estimator, an asymptotically consistent estimate of covariance matrix can be derived free from the distributional assumptions of observations (Hardin & Hilbe, 2007; Huber, 1967; White, 1980).

There is another feasible but rarely mentioned modeling strategy for complex survey data called the *maximum model* (Hox, 2002). In the maximum model strategy, a saturated between-level model is specified; that is, all the unique elements in the between-level variance–covariance matrix are estimated, resulting in the consumption of all available degrees of freedom in the higher level. In other words, we estimate all possible correlations among variables in the between-level.

Originally suggested as the baseline model before any further higher level model construction with theoretical evidence, the maximum model has been discussed by several researchers (e.g., Hox, 2002; Stapleton, 2006; Yuan & Bentler, 2007). In a recent study, Ryu and West (2009) employed the maximum modeling strategy to examine the performance of level-specific fit indexes. Nevertheless, the performance of this modeling strategy for addressing data dependency has not yet been systemically examined.

The purpose of this study is to compare potential differences on the overall model chi-square test and several commonly used fit indexes, the parameter estimates, 95% coverage for both fixed-effect and random-effect estimates, and the respective statistical inferences when analyzing multilevel data with a design-based single-level confirmatory factor analytic (CFA) model and two model-based multilevel CFA (MCFA) models (i.e., the two-level true model and the maximum model). Specifically, our major research question is this: What are the effects of the number of clusters, cluster size, ICC, and model specification on the overall model fit indexes, the fixed-effect and random-effect estimates, the 95% coverage rate, and respective statistical inferences when:

1. The between-level and within-level have the same model structure.
2. The between-level and within-level have different model structures, including
   a. Complex within-level and simple between-level structure.
   b. Simple within-level and complex between-level structure?

METHOD

Three simulation scenarios were created to answer these research questions. We had equal between-level and within-level structures for Scenario 1, a simple between-level and complex within-level structures for Scenario 2, and a complex between-level and simple within-level structures for Scenario 3. The details of these scenarios are described later as well as in both Table 1 and Figure 1. In each scenario, a two-level MCFA model was constructed. This MCFA model was similar to the model examined by Yuan and Bentler (2002) with some modifications referred to Hu and Bentler's (1998, 1999) research. Four factors were controlled when conducting each scenario simulation: cluster number (CN = 50, 150, and 300; Hox & Maas, 2001; B. O. Muthén & Satorra, 1995), cluster size (CS = 10, 50, and 200; Hox & Maas, 2001), ICC (low ICC = 0.1 and high ICC = 0.5; Hox & Maas, 2001), and model specification. The Monte Carlo procedure of M*plus* 6.0 (L. K. Muthén & Muthén, 1998–2010) was used to produce 1,000 replications for each combination of factors in each scenario; that is, a total of: 3 (scenarios) × 3 (cluster numbers) × 3 (cluster sizes) × 2 (ICCs) × 1,000 = 54,000 replications were generated.

TABLE 1
The Data Generation Scheme for the Three Scenarios

	Scenario		
	1	*2*	*3*
Within-level structure	(A) Complex 3-factor	(A) Complex 3-factor	(B) Simple 1-factor
Between-level structure	(A) Complex 3-factor	(B) Simple 1-factor	(A) Complex 3-factor

Note. Scenario 1: Equal between-level model/within-level model. Scenario 2: Simple between-level model/complex within-level model. Scenario 3: Complex between-level model/simple within-level model.

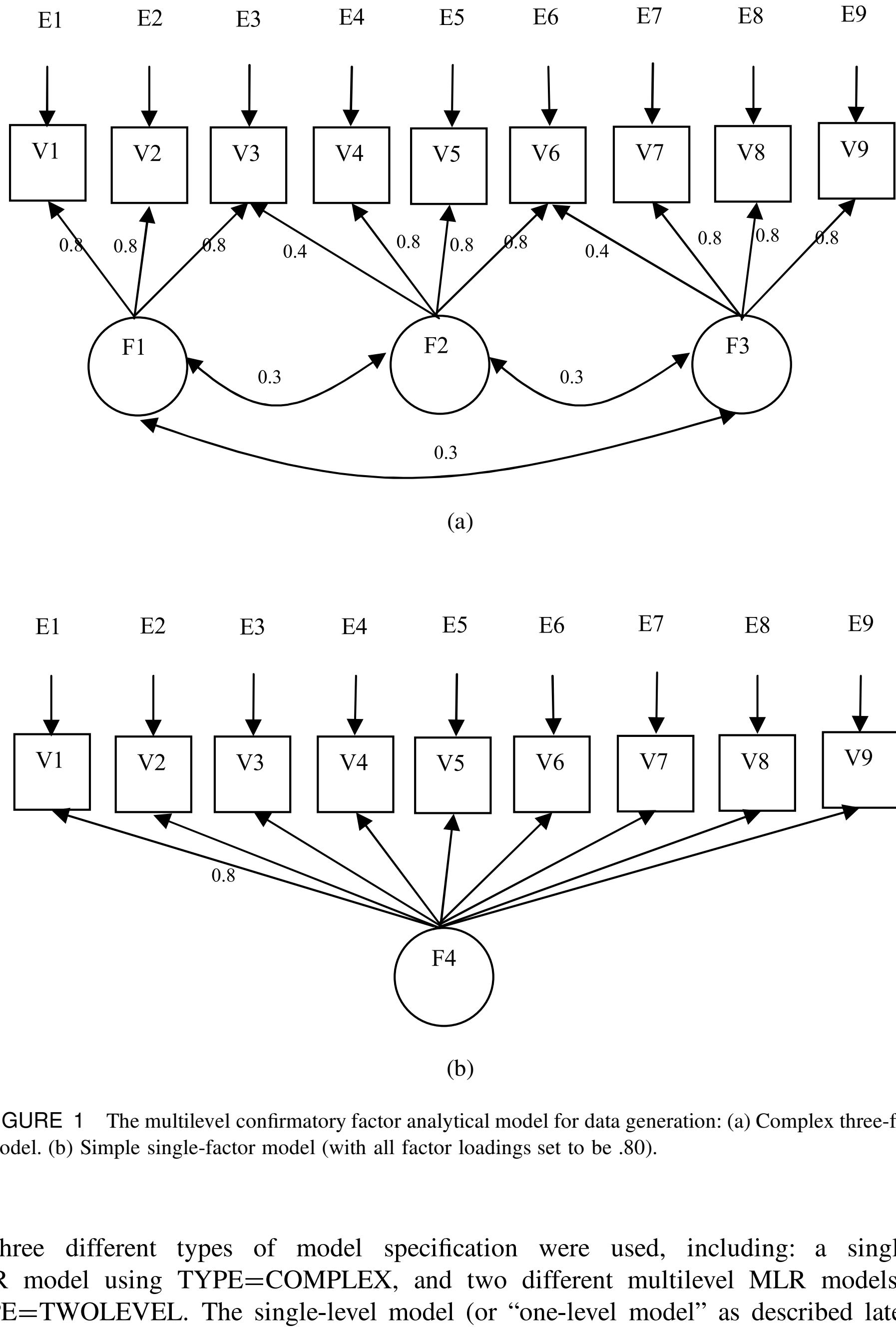

FIGURE 1 The multilevel confirmatory factor analytical model for data generation: (a) Complex three-factor model. (b) Simple single-factor model (with all factor loadings set to be .80).

Three different types of model specification were used, including: a single-level MLR model using TYPE=COMPLEX, and two different multilevel MLR models using TYPE=TWOLEVEL. The single-level model (or "one-level model" as described later) had correctly specified the within-level structure without modeling the between-level structure. Both of the two multilevel MLR models also had correctly specified the within structure. One model had a correctly specified between-level structure (i.e., the "two-level true model"), and the other had a saturated between-level structure (i.e., the "two-level maximum model" with all the nondirectional parameters in the between-covariance matrix being free for estimation). All generated data sets (i.e., the 54,000 replications) were analyzed separately by these three model specifications using M*plus* 6.0 (L. K. Muthén & Muthén, 1998–2010). Detailed information of each scenario was depicted next as well as in Table 1 and Figure 1.

Scenario 1: Equal Between-Level Model and Within-Level Model

As presented in Table 1, the complex survey data were generated based on an equal between- and within-level three-factor model with nine observed variables loaded on three common factors (i.e., Figure 1a). By following previously published simulation studies (e.g., Hox & Maas, 2001; B. O. Muthén & Satorra, 1995), the correlations between the common factors were set to .3, and most of the factor loadings between the latent factors and the outcomes were assigned to be .8. Two cross-loaded factor loadings were specified to .4 in both within-level and between-level models (i.e., F2 $\rightarrow$ V3 and F3 $\rightarrow$ V6). The residual variances of all outcome variables were taken as values that would mostly yield unit-variance measured variables under normality (Hu & Bentler, 1998) and were specified to be equal to .36.

Scenario 2: Simple Between-Level Model/Complex Within-Level Model

In Scenario 2, the between-level model was reduced to a single-factor model as shown in Figure 1b, and the within-level model was the same as the within-level model in Scenario 1 and Figure 1a. The parameterization of the within-level model in Scenario 2 was exactly the same as the within-level model in Scenario 1. In the between-level model, all factor loadings were fixed at .8 and the residual variances of the outcome variables were set to .36.

Scenario 3: Complex Between-Level Model/Simple Within-Level Model

The model setup in Scenario 3 was the opposite of Scenario 2. In Scenario 3, the between-level model was a three-factor model as shown in Figure 1a, and the within-level model was a single-factor model as shown in Figure 1b. The parameterization of the between- and within-level models in Scenario 3 was the same as the corresponding factor models specified in Scenario 2.

RESULTS

The results for each scenario were presented in the following order: the convergence rate, overall model chi-square test statistic and fit indexes, fixed effect estimates, and random effect estimates for each scenario, respectively.

Scenario 1: Equal Between-Level Model and Within-Level Model

In general, all three modeling approaches resulted in adequate model fit and similar parameter estimates that were very close to the true parameter values. We have summarized the findings here; complete results for Scenario 1 can be obtained from the first author on request. All replications converged for each of the three modeling approaches (i.e., the one-level model, the two-level true model, and the maximum model).

Evaluation of test statistic and model fit indexes. Both the one-level model and the two-level maximum model chi-square values were generally close to the theoretical value (i.e., 22), which was the degrees of freedom of the two models across all simulation conditions. Similarly, the chi-square values of the two-level true model were close to the theoretical value (i.e., 44) under the large cluster size condition (CN = 300) but inflated under the smaller cluster size condition (e.g., [CN, CS, ICC] = [50, 10, 0.5], $\chi^2 = 50.28$).

According to recommended cutoff criteria for commonly used fit indexes (e.g., comparative fit index [CFI] $>$ .95, root mean squared error of approximation [RMSEA] $\leq$ .05, and standardized root mean square residual [SRMR] $\leq$ .05 for good fit models), all three modeling approaches showed adequate model fit to the dependent data with equal between/within structure. That is, all models had a CFI greater than 0.99 and an RMSEA smaller than 0.019. Two different SRMRs, SRMR-between and SRMR-within, were reported for the two-level true model and the two-level maximum model. Only one SRMR was reported in the one-level model. Most SRMRs were smaller than the recommended cutoff value (.05). Only a few of the SRMR-between of the two-level true model were larger than the recommended cutoff value, which were all from the small number of clusters condition (i.e., CN = 50).

Fixed effect estimates. In general, one-level model produced consistent and efficient estimates of factor loadings in Scenario 1. As for the two-level true model, the parameter estimates in the within-level model were more consistent and efficient than the parameter estimates in the between-level model. The between-level model parameter estimates were less consistent and efficient because of the larger sample size in the within-level model and the relatively smaller sample size (i.e., number of clusters) in the between-level model. Moreover, in the high ICC setting, the between-level model produced more efficient fixed effect estimates than in the low ICC setting. When the two-level maximum model was used, the fixed effect estimates were more consistent and efficient. In addition, the standard errors were smaller than those from the one-level model and closer to the two-level true model values. All three modeling approaches mostly resulted in statistically significant fixed effect estimates. Only in low ICC and small overall sample size conditions some between-level fixed effect estimates of the two-level true model were not statistically significant.

Random effect estimates. All three modeling approaches resulted in similar estimate patterns for both residual variances and factor variances. In general, all estimates were consistent and became more efficient (with smaller standard error) as the overall sample size increased. For the factor covariances, the two-level true model and the two-level maximum model had consistent estimates around the population value (.30). The factor covariance estimates from the one-level model were around .600, twice the size of the factor covariance estimates in the two-level true model. This enlarged estimate was mainly due to the redistribution of the one-level model factor variance. Factor covariances were mostly significant in all three modeling approaches. Only some between-level factor covariances in the two-level true model were not statistically significant, generally under the small number of clusters condition.

Scenario 2: Simple Between-Level Model/Complex Within-Level Model

The convergence rate was generally equal to 100% for all three modeling approaches. A few of the replications with low ICC and small overall sample size conditions for the two-level true model were not equal to 100% (e.g., convergence rate reduced from 99% with CS = 50 to 95% with CS = 10).

Evaluation of test statistic and model fit indexes. Model fit indexes showed adequate fit in all three modeling approaches even though the one-level model assumed equivalent between- and within-model structure. Table 2 presents the overall model-fit chi-square test statistic (shown as χ^2), CFI, RMSEA, and SRMR for the three modeling approaches under different simulation settings. Due to the unequal between- and within-level structures, chi-square values in the one-level model deviated from the theoretical chi-square value (i.e., model $df = 22$). However, the two-level maximum model ($df = 22$) and the two-level true model ($df = 49$) still had chi-square values close to their degrees of freedom across all settings. Both CFI and RMSEA for the one-level model still showed very good fit even though the between- and within-models were different from each other. SRMR-between and SRMR-within were reported for the two-level true and maximum models and a single SRMR was reported for the one-level model. Similarly, most of the SRMRs, especially the ones from the one-level model, indicated good model fit.

Fixed effect estimates. For conciseness, only the fixed effect estimates of F2 by V3, V4, V5, and V6 in the within-level model were reported because they contained both single-loaded (i.e., V4 and V5) and cross-loaded (i.e., V3 and V6) observed indicators. The estimates and standard errors of the other single-loaded factor loadings and the corresponding residual variances of the observed indicators V1, V2, V7, V8, and V9 were very similar to those of V4 and V5. The between-level factor loading estimates of F4 by V3, V4, V5, and V6 in the two-level true model were reported for illustration purposes. According to Table 3, the two-level true model produced unbiased factor loading estimates for both single- and cross-loaded variables. However, the one-level model in the large sample size setting yielded unbiased single-loaded factor loading estimates but inconsistent cross-loaded factor loading estimates. For example, the parameter estimates for the cross-loaded observed variables, V3 and V6, were substantially underestimated ($\hat{\lambda}_{Y3FW2} = 0.290$ and $\hat{\lambda}_{Y6FW2} = 0.679$ given [CN, CS, ICC] = [50, 10, 0.5]).

The standard errors of the fixed effect estimates also exhibited distinct patterns. In the two-level true model, the standard errors of the factor loadings were larger in the between-level model than those in the within-level model. For the one-level model, the standard errors of the loading estimates became smaller as ICCs decreased. Nevertheless, the 95% coverage rate of the cross-loaded fixed effects (i.e., from F2 to V3 and V6) was very poor in the one-level model, mainly due to the seriously attenuated factor loading estimates (e.g., given [CN, CS, ICC] = [300, 200, 0.5], 95% coverage rate = 0.023 for V6 and 0.013 for V3 in the one-level model). Conversely, the two-level maximum model produced unbiased and efficient factor loading estimates for both single-loaded and cross-loaded indicators.

Random effect estimates. We decided to only present the result of random effects under the large sample size condition (i.e., [CN, CS] = [200, 300]) given that the residual variances,

TABLE 2
Test Statistic and Model Fit Indexes for Scenario 2: Simple Between-Level/Complex Within-Level Structures

				SRMR	
Model	χ^2	*CFI*	*RMSEA*	*Between*	*Within*
ICC = .1					
50(10)					
Two-level	58.716	.998	.017	.098	.010
One-level	30.666	.995	.024	—	.015
Maximum model	23.774	.999	.012	.006	.017
50(200)					
Two-level	54.109	1.000	.003	.043	.002
One-level	34.639	.995	.007	—	.011
Maximum model	22.656	1.000	.002	.000	.002
300(10)					
Two-level	51.136	1.000	.004	.045	.004
One-level	25.156	.999	.006	—	.011
Maximum model	22.243	1.000	.004	.002	.007
300(200)					
Two-level	49.416	1.000	.001	.017	.001
One-level	25.483	.999	.001	—	.009
Maximum model	22.216	1.000	.001	.000	.001
ICC = .5					
50(10)					
Two-level	55.549	.996	.014	.050	.020
One-level	28.069	.993	.020	—	.024
Maximum model	23.250	.999	.011	.002	.020
50(200)					
Two-level	53.987	1.000	.003	.041	.004
One-level	29.733	.995	.005	—	.020
Maximum model	22.703	1.000	.002	.000	.004
300(10)					
Two-level	50.571	1.000	.004	.020	.008
One-level	33.763	.998	.012	—	.012
Maximum model	22.248	1.000	.004	.001	.008
300(200)					
Two-level	49.453	1.000	.001	.016	.002
One-level	37.626	.998	.003	—	.011
Maximum model	22.226	1.000	.001	.000	.002

Note. Two-level = two-level true model ($df = 49$); one-level = one-level model ($df = 22$); maximum model = two-level maximum model ($df = 22$). In the model column, 50(10) represents cluster number = 50 and cluster size = 10; thus, sample size for this setting equals 500. The same notation is used for the rest of the settings. χ^2 = overall model chi-square test statistics; ICC = intraclass correlation; CFI = comparative fit index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square residual.

factor variances, and covariances had exactly the same pattern of results across all sample size settings in all three modeling strategies. The complete results can be obtained from the first author on request. As shown in Table 4, the estimated factor variances of the one-level model were twice the size of those from the two-level true and maximum models. The two two-level models produced almost identical results. As for residual variance, the one-level model produced single-loaded indicator residual variances close to the sum of the residual

TABLE 3
Fixed Effects Estimates of Scenario 2: Simple Between-Level/Complex Within-Level Structures

	Two-Level True Model						*One-Level Model*			*Two-Level Maximum Model*		
	Estimates		*95% Cover*		*% Significant Coefficient*							
Fixed Effect	*Between*	*Within*	*Between*	*Within*	*Between*	*Within*	*Estimates*	*95% Cover*	*% Significant Coefficient*	*Estimates*	*95% Cover*	*% Significant Coefficient*
ICC = .1												
50(10)												
F4/F2 by												
V4	1.028 (1.153)	0.800 (.000)	0.937	1.000	0.235	0.000	0.800 (.000)	1.000	0.000	0.800 (.000)	1.000	0.000
V5	1.074 (1.907)	0.798 (.034)	0.929	0.940	0.249	0.996	0.799 (.020)	0.959	1.000	0.799 (.033)	0.944	1.000
V6	1.008 (1.312)	0.800 (.036)	0.929	0.935	0.237	0.996	0.768 (.023)	0.686	1.000	0.801 (.034)	0.938	1.000
V3	1.091 (3.788)	0.399 (.030)	0.939	0.956	0.300	0.996	0.370 (.019)	0.668	1.000	0.401 (.030)	0.935	1.000
300(200)												
F4/F2 by												
V4	0.797 (.053)	0.800 (.000)	0.954	1.000	1.000	0.000	0.800 (.000)	1.000	0.000	0.800 (.000)	1.000	0.000
V5	0.798 (.054)	0.800 (.001)	0.949	0.948	1.000	1.000	0.799 (.018)	.957	1.000	0.800 (.003)	0.945	1.000
V6	0.799 (.054)	0.799 (.001)	0.950	0.951	1.000	1.000	0.769 (.019)	.629	1.000	0.800 (.003)	0.950	1.000
V3	0.800 (.052)	0.400 (.001)	0.945	0.957	1.000	1.000	0.372 (.014)	.500	1.000	0.400 (.002)	0.948	1.000
ICC = .5												
50(10)												
F4/F2 by												
V4	0.818 (.152)	0.800 (.000)	0.948	1.000	1.000	0.000	0.800 (.000)	1.000	0.000	0.800 (.000)	1.000	0.000
V5	0.819 (.153)	0.799 (.048)	0.938	0.944	1.000	1.000	0.804 (.081)	0.943	1.000	0.800 (.047)	0.937	1.000
V6	0.819 (.153)	0.802 (.051)	0.937	0.928	1.000	1.000	0.679 (.090)	0.680	1.000	0.803 (.051)	0.933	1.000
V3	0.818 (.145)	0.399 (.044)	0.926	0.930	0.999	1.000	0.290 (.079)	0.688	0.948	0.400 (.044)	0.928	1.000
300(200)												
F4/F2 by												
V4	0.797 (.051)	0.800 (.000)	0.943	1.000	1.000	0.000	0.800 (.000)	1.000	0.000	0.800 (.000)	1.000	0.000
V5	0.798 (.051)	0.800 (.004)	0.941	0.949	1.000	1.000	0.801 (.029)	0.945	1.000	0.800 (.004)	0.949	1.000
V6	0.799 (.051)	0.799 (.005)	0.939	0.952	1.000	1.000	0.672 (.029)	0.023	1.000	0.800 (.005)	0.952	1.000
V3	0.800 (.051)	0.399 (.004)	0.944	0.942	1.000	1.000	0.289 (.025)	0.013	1.000	0.400 (.004)	0.942	1.000

Note. Standard error of parameter estimates is shown in parentheses. 95% Cover = proportion of replications for which the 95% confidence interval contained the true population parameter value; % Significant Coefficient = proportion of replications that produced statistically significant estimates; ICC = intraclass correlation. The factor loadings between the between-level factor F4 of V4, V5, V6, and V3were fixed at 0.8. The factor loadings between the within-level factor F2 and V4, V5, and V6 were fixed at 0.8, and V3 was set at 0.4 in the within-level model. For marker variables, their 95% Cover was reported as 1 and % Significant Coefficient as 0.

TABLE 4
Random Effect Estimates of Large Sample Size Setting in Scenario 2: Simple Between-Level/Complex Within-Level Structures

	Two-Level True Model						*One-Level Model*			*Two-Level Maximum Model*		
	Estimates		*95% Cover*		*% Significant Coefficient*							
Random Effects	*Between*	*Within*	*Between*	*Within*	*Between*	*Within*	*Estimates*	*95% Cover*	*% Significant Coefficient*	*Estimates*	*95% Cover*	*% Significant Coefficient*
ICC = .1												
Factor variance												
F4/F1	0.207 (.050)	1.799 (.014)	0.940	0.948	1.000	1.000	2.005 (.070)	0.164	1.000	1.799 (.014)	0.948	1.000
F2		1.800 (.014)		0.956		1.000	1.999 (.065)	0.129	1.000	1.800 (.014)	0.956	1.000
F3		1.800 (.014)		0.941		1.000	1.999 (.068)	0.163	1.000	1.800 (.014)	0.941	1.000
Factor covariance												
F1 with F2		0.300 (.008)		0.946		1.000	0.481 (.030)	0.000	1.000	0.300 (.008)	0.947	1.000
F1 with F3		0.299 (.008)		0.941		1.000	0.477 (.028)	0.000	1.000	0.299 (.008)	0.937	1.000
F2 with F3		0.300 (.008)		0.970		1.000	0.490 (.030)	0.000	1.000	0.300 (.008)	0.934	1.000
Residual variance												
V3	0.357 (.033)	0.360 (.003)	0.933	0.955	1.000	1.000	0.744 (.048)	0.000	1.000	0.360 (.003)	0.955	1.000
V4	0.359 (.033)	0.359 (.003)	0.951	0.951	1.000	1.000	0.718 (.044)	0.000	1.000	0.359 (.003)	0.951	1.000
V5	0.358 (.033)	0.360 (.003)	0.944	0.953	1.000	1.000	0.720 (.044)	0.000	1.000	0.360 (.003)	0.953	1.000
V6	0.357 (.033)	0.360 (.003)	0.947	0.953	1.000	1.000	0.743 (.046)	0.000	1.000	0.360 (.003)	0.953	1.000
ICC = .5												
Factor variance												
F4/F1	1.009 (.122)	0.999 (.009)	0.943	0.938	1.000	1.000	2.021 (.132)	0.000	1.000	0.999 (.009)	0.937	1.000
F2		1.000 (.009)		0.962		1.000	2.010 (.129)	0.000	1.000	1.000 (.009)	0.962	1.000
F3		1.000 (.009)		0.943		1.000	2.003 (.131)	0.000	1.000	1.000 (.009)	0.943	1.000
Factor covariance												
F1 with F2		0.300 (.005)		0.956		1.000	1.277 (.098)	0.000	1.000	0.300 (.005)	0.956	1.000
F1 with F3		0.299 (.005)		0.937		1.000	1.268 (.097)	0.000	1.000	0.299 (.005)	0.937	1.000
F2 with F3		0.300 (.005)		0.935		1.000	1.289 (.098)	0.000	1.000	0.300 (.005)	0.935	1.000
Residual Variance												
V3	0.357(.033)	0.359(.003)	0.930	0.950	1.000	1.000	0.795(.042)	0.000	1.000	0.359 (.003)	0.950	1.000
V4	0.358(.033)	0.359(.003)	0.948	0.958	1.000	1.000	0.711(.039)	0.000	1.000	0.359 (.003)	0.958	1.000
V5	0.359(.033)	0.360(.003)	0.939	0.950	1.000	1.000	0.712(.040)	0.000	1.000	0.360 (.003)	0.950	1.000
V6	0.358(.033)	0.360(.003)	0.951	0.951	1.000	1.000	0.790(.041)	0.000	1.000	0.360 (.003)	0.951	1.000

Note. Standard error of parameter estimates are shown in parentheses. 95% Cover = proportion of replications for which the 95% confidence interval contained the true population parameter value; % Significant Coefficient = proportion of replications that produced statistically significant estimates; ICC = intraclass correlation. Factor variance was set as 0.2 in between-level model and as 1.8 in within-level model in low ICC setting and was set as 1 in both levels in high ICC setting. The factor covariance was set at 0.3, and the residual variance was set at 0.36.

variance estimates in the between- and within-levels of the two-level true model. However, for the residual variance of the cross-loaded indicators (i.e., V3 & V6), particularly under the high ICC condition, the one-level model resulted in overestimated residual variances. The inflated residual variance of cross-loaded indicators in the one-level model was likely due to the inaccurate estimates of the factor loadings. The one-level model also resulted in substantially biased estimates of the factor covariances as shown in the 95% CI coverage rate (ranged from 0–16%).

Scenario 3: Complex Between-Level Model/Simple Within-Level Model

Most of the analyses for the three modeling approaches converged. Some two-level true model replications did not converge, especially those with small overall sample size (e.g., given the number of clusters equal to 50, the convergence rate reduced from 96% for cluster size = 50 to 89% for cluster size = 10).

Evaluation of test statistic and model fit indexes. Unlike the previous two scenarios, the model fit chi-square test statistic and some of the fit indexes showed noticeable differences between the one-level model and the two two-level models. Table 5 compared overall model-fit chi-square test statistic (shown as χ^2), CFI, RMSEA, and SRMR for the three modeling approaches. The overall model chi-square test statistic of the one-level model deviated from the theoretical value (i.e., the model $df = 27$). As sample size increased, the discrepancy became larger.

For the model fit indexes, all three modeling approaches resulted in fairly good model fit under the low ICC condition. However, as ICC increased to 0.5, the one-level model fit indexes showed incongruent patterns of lack of fit. CFIs consistently showed lack of fit (ranging from .74–.78), signaling unequal between- and within-structures. Only some of the one-level model RMSEAs (especially the ones under the small cluster size condition) indicated different model structure at a different level. Unlike CFI and RMSEA, most of the SRMRs still exhibited good fit of the one-level model to the data.

Fixed effect estimates. Fixed effect estimates of the factor loadings between indicators (V3, V4, V5, and V6) and the between-level factor (F2) as well as the within-level factor (F4) were tabulated in Table 6. In the small sample size and low ICC settings, the two-level true model generated inconsistent and inefficient estimates of factor loadings for both single- and cross-loaded indicators, except for V4, which was the marker variable (e.g., $\hat{\lambda}_{V5F2} = 1.085$, $\hat{\lambda}_{V6F2} = 1.015$, and $\hat{\lambda}_{V3F2} = 0.372$ at [CN, CS, ICC] = [50, 10, 0.1]). The one-level model generated biased fixed effect estimates for cross-loaded indicators (V3 and V6) that positively deviated (or overestimated) from the population values. The proportion of the bias in the estimate to the true parameter value ranged from 6.3% to 29% as ICC increased from 0.1 to 0.5. Moreover, the factor loading estimates of the single-loaded indicators (V4 and V5) in the one-level model were also overestimated with relative bias as large as 8% more than the true parameter value. The standard errors of the fixed effect estimates also exhibited inflated patterns as ICC increased in the one-level model. Conversely, the two-level maximum model generated unbiased estimates of the fixed effects and standard errors for both single- and cross-loaded indicators in the within-level model.

TABLE 5
Test Statistic and Model Fit Indexes for Scenario 3: Complex Between-Level/Simple Within-Level Structures

				SRMR	
Model	χ^2	*CFI*	*RMSEA*	*Between*	*Within*
ICC = .1					
50(10)					
Two-level	55.830	1.000	.021	.099	.019
One-level	44.471	.991	.030	—	.026
Maximum model	28.626	1.000	.019	.008	.018
50(200)					
Two-level	53.981	1.000	.003	.074	.003
One-level	49.748	.992	.017	—	.021
Maximum model	27.932	1.000	.003	.000	.003
300(10)					
Two-level	49.861	1.000	.004	.038	.009
One-level	95.882	.997	.033	—	.019
Maximum model	27.024	1.000	.003	.003	.008
300(200)					
Two-level	50.014	1.000	.001	.030	.001
One-level	120.653	.996	.010	—	.015
Maximum model	27.281	1.000	.001	.000	.001
ICC = .5					
50(10)					
Two-level	55.680	1.000	.018	.076	.013
One-level	187.072	.762	.112	—	.089
Maximum model	28.615	1.000	.017	.003	.012
50(200)					
Two-level	54.216	1.000	.002	.062	.002
One-level	238.213	.744	.035	—	.083
Maximum model	27.904	1.000	.001	.000	.002
300(10)					
Two-level	49.694	1.000	.002	.032	.010
One-level	816.442	.781	.108	—	.074
Maximum model	27.031	1.000	.001	.001	.010
300(200)					
Two-level	49.912	1.000	.001	.020	.002
One-level	1021.833	.772	.034	—	.072
Maximum model	27.271	1.000	.001	.000	.002

Note. Two-level = two-level true model (*df* = 49); one-level = one-level model (*df* = 27); maximum model = two-level maximum model (*df* = 27). In the model column, 50(10) represents cluster number = 50 and cluster size = 10; thus, sample size for this setting equals 500. The same notation is used for the rest of the settings. χ^2 = overall model chi-Square test statistics; ICC = intraclass correlation; CFI = comparative fit index; RMSEA = root mean squared error of approximation; SRMR = standardized root mean square residual.

Random effect estimates. Results of the random effect estimates from the largest sample size (CS = 300, CN = 200) condition are reported in Table 7. The two-level maximum model produced almost identical within-level random effect estimate results as the two-level true model, including the variance component estimates with standard errors and the corresponding statistical inferences. On the other hand, both factor and residual variance estimates in the one-level model substantially deviated from the true parameter values. When ICC increased from .1

TABLE 6
Fixed Effects Estimates of Scenario 3: Complex Between-Level/Simple Within-Level Structures

	Two-Level True Model						*One-Level Model*			*Two-Level Maximum Model*		
	Estimates		*95% Cover*		*% Significant Coefficient*							
Fixed Effect	*Between*	*Within*	*Between*	*Within*	*Between*	*Within*	*Estimates*	*95% Cover*	*% Significant Coefficient*	*Estimates*	*95% Cover*	*% Significant Coefficient*
ICC = .1												
50(10)												
F2/F4 by												
V4	0.800 (.000)	0.800 (.032)	1.000	0.934	0.000	0.999	0.799 (0.048)	0.942	1.000	0.801 (.030)	0.934	1.000
V5	1.085 (1.635)	0.801 (.031)	0.935	0.935	0.540	0.999	0.800 (.048)	0.934	1.000	0.801 (.030)	0.935	1.000
V6	1.015 (1.580)	0.801 (.031)	0.908	0.950	0.269	0.999	0.853 (.050)	0.831	1.000	0.801 (.030)	0.950	1.000
V3	0.372 (1.466)	0.801 (.032)	0.974	0.942	0.140	0.999	0.855 (.053)	0.831	1.000	0.801 (.030)	0.942	1.000
300(200)												
F2/F4 by												
V4	0.800 (.000)	0.800 (.003)	1.000	0.949	0.000	1.000	0.798 (.014)	0.936	1.000	0.800 (.003)	0.949	1.000
V5	0.806 (.100)	0.800 (.003)	0.941	0.955	1.000	1.000	0.797 (.014)	0.945	1.000	0.800 (.003)	0.955	1.000
V6	0.812 (.146)	0.800 (.003)	0.937	0.938	1.000	1.000	0.850 (.014)	0.051	1.000	0.800 (.003)	0.938	1.000
V3	0.394 (.169)	0.800 (.003)	0.956	0.942	0.687	1.000	0.853 (.016)	0.066	1.000	0.800 (.003)	0.942	1.000
ICC = .5												
50(10)												
F2/F4 by												
V4	0.800 (.000)	0.801 (.041)	1.000	0.941	0.000	1.000	0.864 (.143)	0.955	1.000	0.801 (.041)	0.939	1.000
V5	0.825 (.182)	0.801 (.041)	0.947	0.939	0.989	1.000	0.865 (.144)	0.954	1.000	0.801 (.041)	0.939	1.000
V6	0.841 (.210)	0.802 (.041)	0.958	0.943	0.976	1.000	1.032 (.167)	0.812	1.000	0.802 (.041)	0.943	1.000
V3	0.413 (.177)	0.801 (.041)	0.959	0.945	0.718	1.000	0.977 (.101)	0.606	1.000	0.801 (.041)	0.943	1.000
300(200)												
F2/F4 by												
V4	0.800 (.000)	0.800 (.004)	1.000	0.952	0.000	1.000	0.849 (.052)	0.893	1.000	0.800 (.004)	0.952	1.000
V5	0.804 (.060)	0.800 (.004)	0.945	0.949	1.000	1.000	0.849 (.052)	0.884	1.000	0.800 (.004)	0.949	1.000
V6	0.804 (.064)	0.800 (.004)	0.947	0.935	1.000	1.000	1.012 (.059)	0.022	1.000	0.800 (.004)	0.935	1.000
V3	0.399 (.055)	0.800 (.004)	0.945	0.938	1.000	1.000	0.968 (.035)	0.000	1.000	0.800 (.004)	0.938	1.000

Note. Standard error of parameter estimates are shown in parentheses. 95% Cover = proportion of replications for which the 95% confidence interval contained the true population parameter value; % Significant Coefficient = proportion of replications that produced statistically significant estimates; ICC = intraclass correlation. The factor loadings between the between-level factor F2 and V4, V5, and V6 were fixed at 0.8, and V3 was set at 0.4. The factor loadings between the within-level factor F4 and V4, V5, V6, and V3 were fixed at 0.8 in the within-level model. For marker variables, their 95% Cover was reported as 1 and % Significant Coefficient as 0.

TABLE 7
Random Effect Estimates of Large Sample Size Setting in Scenario 3: Complex Between-Level/Simple Within-Level Structures

	Two-Level True Model						*One-Level Model*			*Two-Level Maximum Model*		
	Estimates		*95% Cover*		*% Significant Coefficient*							
Random Effects	*Between*	*Within*	*Between*	*Within*	*Between*	*Within*	*Estimates*	*95% Cover*	*% Significant Coefficient*	*Estimates*	*95% Cover*	*% Significant Coefficient*
ICC = .1												
Factor variance												
F1/F4	0.200 (.048)	1.800 (.013)	0.945	0.943	0.997	1.000	2.082 (.058)	0.002	1.000	1.800 (.013)	0.943	1.000
F2	0.200 (.047)		0.939		0.999							
F3	0.200 (.043)		0.943		1.000							
Residual variance												
V4	0.357 (.036)	0.360 (.002)	0.939	0.950	1.000	1.000	0.705 (.031)	0.000	1.000	0.360 (.002)	0.950	1.000
V5	0.359 (.032)	0.360 (.002)	0.940	0.953	1.000	1.000	0.675 (.029)	0.000	1.000	0.360 (.002)	0.953	1.000
V6	0.359 (.032)	0.360 (.002)	0.937	0.952	1.000	1.000	0.674 (.029)	0.000	1.000	0.360 (.002)	0.952	1.000
V3	0.359 (.034)	0.360 (.002)	0.955	0.938	1.000	1.000	0.715 (.032)	0.000	1.000	0.360 (.002)	0.938	1.000
ICC = .5												
Factor variance												
F1/F4	1.000 (.131)	1.000 (.009)	0.945	0.946	1.000	1.000	1.379 (.120)	0.086	1.000	1.000 (.009)	0.945	1.000
F2	0.999 (.129)		0.935		1.000							
F3	0.992 (.128)		0.949		1.000							
Residual variance												
V4	0.354 (.048)	0.360 (.002)	0.933	0.947	1.000	1.000	1.058 (.069)	0.000	1.000	0.360 (.002)	0.947	1.000
V5	0.357 (.044)	0.360 (.002)	0.940	0.954	1.000	1.000	1.007 (.063)	0.000	1.000	0.360 (.002)	0.954	1.000
V6	0.357 (.044)	0.360 (.002)	0.951	0.951	1.000	1.000	1.009 (.063)	0.000	1.000	0.360 (.002)	0.951	1.000
V3	0.355 (.046)	0.360 (.002)	0.941	0.937	1.000	1.000	0.939 (.056)	0.000	1.000	0.360 (.002)	0.937	1.000

Note. Standard error of parameter estimates are shown in parentheses. 95% Cover = proportion of replications for which the 95% confidence interval contained the true population parameter value; % Significant Coefficient = proportion of replications that produced statistically significant estimates; ICC = intraclass correlation. Factor variance was set as 0.2 in between-level model and as 1.8 in within-level model in low ICC setting and was set as 1 in both levels in high ICC setting. Residual variance was set at 0.36.

to .5, the factor variance estimates of the one-level model were underestimated and less efficient (with larger standard error). The relative bias was as large as −31% for ICC = .5. When ICC increased from .1 to .5, the residual variance estimates of both cross-loaded and single-loaded indicators of the one-level model were overestimated and less efficient with relative bias as large as 32%. All the 95% coverage rates of all the factor and residual variance estimates in the one-level model were close to zero due to the substantially biased point estimates.

DISCUSSION

One interesting finding from this simulation study is that the overall model chi-square test and commonly used fit indexes could not consistently provide much helpful information on the necessity of specifying a different higher level model. This occurred more often when the design-based single-level model was used for analyzing data with truly unequal between-level and within-level model structures. For example, although the model fit test statistic values for the one-level model deviated slightly from the expected values ($df = 22$) due to the unequal between and within structures, the p value of the overall model chi-square value in the simple between-/complex within-level structure scenario (i.e., Scenario 2), was still larger than .05. Overall model chi-square values that are not statistically significant can lead to erroneous conclusions about the equivalence of the between-level and within-level models. Similarly, based on the traditional cutoff criteria of three commonly used fit indexes (i.e., RMSEA < .08, SRMR < .08, and CFI > .95 for adequately fitted models), none of these three fit indexes from the one-level model was sensitive to deviations between the within-level model and the between-level model.

On the other hand, in Scenario 3 (i.e., complex-between/simple-within structure), the overall model chi-square test of the design-based one-level model indicated a poor fit. As for the model fit indexes, only the CFI of the one-level model consistently showed poor model fit under the high ICC condition (i.e., ICC = 0.5). In general, both RMSEA and SRMR failed to indicate the lack of fit of the one-level model to the multilevel data with different model structures at different levels. To sum up, the overall model chi-square test statistic and CFI model fit index can only offer partial information regarding model misspecification when the design-based approach is used for analyzing the clustered data.

For the fixed effects estimates, our simulation results showed different patterns for the equal-structure and unequal-structure scenarios. In the equal-structure scenario, all factor loading estimates were close to the population values in all three modeling approaches. However, in the unequal-structure scenario, different patterns of result occurred for the single- and cross-loaded factor loadings. In the simple between-/complex within-level scenario (Scenario 2), the estimates of the single-loaded factor loadings were generally unbiased. The statistical inferences of these single-loaded factor loadings in the one-level model were close to those of the within-level model in both two-level true and maximum models. Nevertheless, the estimates of the cross-loaded indicators of the one-level model were biased. These biased cross-loaded factor loading estimates were exacerbated as the ICC increased and the standard errors became larger. The corresponding 95% confidence interval coverage rates for these loading estimates were generally low due to the substantially biased factor loading estimates. The situation grew worse when the between-level structure became more complex than the within-level structure (i.e., Scenario 3: the complex between-/simple within-level scenario). The factor loading estimates of both single-loaded and cross-loaded indicators in the one-level model seriously deviated from the population values. Biases became more severe as ICC increased. Additionally, almost all 95% confidence interval coverage rates of biased cross-loading estimate in the one-level model were close to zero, especially under the large sample size condition.

For the random effect estimates, the design-based one-level model estimated the overall variance components (i.e., the combination of the between- and within-level variances) in the equal between- and within-structure scenario (i.e., Scenario 1). For the unequal structures with simple between-level and complex within-level models (i.e., Scenario 2), the design-based one-level model could still produce fairly unbiased estimates of the sum of the between- and within-factor variances. However, the estimate of the factor covariance in the one-level model no longer equaled the sum of the factor covariance from each level. Moreover, the estimates of the residual variances of the cross-loaded indicators were substantially biased. A more severe pattern of biased factor and residual variances of the one-level model was found under the complex between-/simple within-level scenario (i.e., Scenario 3).

To understand the change in the estimation of variances and covariances of the random effects, we need to take the variation of the fixed effect estimates into account. In the equal structure scenario, the fixed effect estimates for the one-level model were asymptotically close to the population values. With unbiased fixed effect estimates, the variance and covariance components for the latent factors as well as the residuals were directly the sum of the corresponding between- and within-level variance components for the one-level model.

However, the variance component estimates in the one-level model were no longer the sum of the different level component estimates when these different levels had different model structures. In the simple-between/complex-within scenario (i.e., Scenario 2), the single-loaded factor loading estimates were generally consistent, whereas the cross-loaded estimates gradually deviated from the true parameter values as ICC increased. Because of the use of the marker variable strategy for model identification, the factor variances were defined by the metric of the marker variable (e.g., V4 loaded on F2 in Figure 1a and Table 3) when both factor covariances and residual variances were allowed to vary. To maintain a consistent amount of indicator variance and to compensate for the inaccurate estimates of the cross-loaded factor loadings, the variance and covariance estimates had to be adjusted. As a result, factor covariance estimates were no longer close to the summed between- and within-level factor covariance. Moreover, the unexplained portion of the indicator variance was then added to the residual variance. For this reason the cross-loaded indicators had higher residual variances than the single-loaded indicators. This situation further deteriorated in the complex between-/simple within-level scenario (i.e., Scenario 3). The one-level model produced an inconsistent factor variance estimate due to the biased single-loaded factor loading estimates. As ICC increased, the underestimation of the factor variance became larger because of the inflated single-loaded factor loading estimates.

Conversely, the two-level maximum model generally resulted in unbiased single- and cross-loaded fixed effect estimates in both Scenarios 2 (simple between-/complex within-level scenario) and 3 (complex between-/simple within-level scenario). Thus, we could obtain consistent variance and covariance estimates of the latent factors and residuals that were close to the true variance components in the within-level of the two-level true model. Moreover, the two-level maximum model, compared to the two-level true model, offered greater statistical power for testing the lower level estimates, especially under small sample size conditions. Although the two-level true model correctly specified the multilevel structure of the data and yielded the asymptotically identical lower level estimates as should occur under most conditions, the two-level true model might still produce parameter estimates with inflated standard error, and inflated Type II error rates, especially under conditions with low ICC and small sample size.

CONCLUSIONS AND SUGGESTIONS

Our study compared the similarities and differences of analyzing complex survey data with equal and unequal multilevel structures using a design-based single-level CFA model (the one-level model) and two model-based multilevel CFA models (the two-level true model and the two-level maximum model). In particular, we examined overall model fit indexes, parameter estimates, 95% coverage for both fixed and random effects, and statistical inferences for detecting the parameter estimates.

The simulation showed that the one-level model (the design-based model) provided satisfying results only under equal between/within structures. Under the simple between-/complex within-level structure, the one-level model yielded erroneous cross-loaded factor loading estimates and biased random effect estimates. As the between-level structure became more complicated than the within-level structure (i.e., Scenario 3: complex between-/simple within-level structure), the design-based approach produced biased estimates for both single- and cross-loaded factor loadings as well as for the random effect variances and covariances.

Modeling the data structures using the two model-based multilevel models turned out to be the best analytical strategy for analyzing multilevel data. However, the higher level model structure might not always be the focus or interest of a study in which there is no hypothesized model for the higher level. As indicated previously, this might be the reason that the design-based single-level approach is commonly used for analyzing multilevel data given its simplicity (i.e., only one model is needed for specification). Under such circumstances, constructing multilevel models can be difficult and daunting for researchers with limited higher level information from the available data, theories, or prior research. The two-level maximum model, where the between-level model is a saturated model (i.e., estimating all the unique nondirectional parameters in the between-level model), is a better and more feasible alternative than the design-based one-level approach and the model-based two-level model. Thus, if a researcher's goal is only to validate the pooled-within covariance structure of the complex survey data (in other words, only a within-level model is of interest and hypothesized), we recommend the use of the two-level maximum model instead of the design-based one-level model. In these circumstances, the two-level maximum model generally produces more consistent and efficient model parameter estimates.

ACKNOWLEDGMENT

The authors are grateful to Sandra Acosta and Yuan-Hsuan Lee for their comments on an earlier version of this manuscript.

REFERENCES

Agrawal, A., & Lynskey, M. T. (2007). Does gender contribute to heterogeneity in criteria for cannabis abuse and dependence? Results from the national epidemiological survey on alcohol and related conditions. *Drug and Alcohol Dependence, 88*(2–3), 300–307.

Au, K., & Cheung, M. W. L. (2004). Intra-cultural variation and job autonomy in 42 countries. *Organization Studies, 25*(8), 1339–1362.

Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variable from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods, 12*, 45–57.

Davidov, E., Yang-Hansen, K., Gustafsson, J. E., Schmidt, P., & Bamberg, S. (2006). Does money matter? A theory-driven growth mixture model to explain travel-mode choice with experimental data. *Methodology, 2*, 124–134.

De Leeuw, J., & Kreft, I. G. (1995). Questioning multilevel models. *Journal of Educational and Behavioral Statistics, 20*(2), 171–190.

du Toit, S. H. C., & du Toit, M. (2008). Multilevel structural equation modeling. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 435–478). New York, NY: Springer.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Boston, MA: Allyn & Bacon.

Hardin, J. W., & Hilbe, J. M. (2007). *Generalized linear models and extensions* (2nd ed.). College Station, TX: Stata Press.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23*(6), 723.

Holt, D., Scott, A. J., & Ewings, P. D. (1980). Chi-squared tests with survey data. *Journal of the Royal Statistical Society, Series A (General), 143*, 303–320.

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hox, J., & Kleiboer, A. M. (2007). Retrospective questions or a dairy method? A two-level multitrait-multimethod analysis. *Structural Equation Modeling, 14*, 311–325.

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling, 8*, 157–174.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 221–233). Berkeley, CA: University of California Press.

Kalton, G. (1997). Practical methods for estimating survey sampling errors. *Bulletin of the International Statistical Institute, 47,* 495–514.

Kish, L. (1965). *Survey sampling.* New York, NY: Wiley.

Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. (2001). Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology, 86*(1), 3–16.

Klein, K. J., & Kozlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods, 3*(3), 211–236.

Lee, E. S., Forthofer, R. N., & Lorimor, R. J. (2006). *Analyzing complex survey data.* Newbury Park, CA: Sage.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. O. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effect in contextual studies. *Psychological Methods, 13*, 203–229.

Mathews, C., Aarø, L. E., Flisher, A. J., Mukoma, W., Wubs, A. G., & Schaalma, H. (2009). Predictors of early first sexual intercourse among adolescents in Cape Town, South Africa. *Health Education Research, 24*, 1–10.

Muthén, B. O., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors, 31*, 1050–1066.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25,* 267–316.

Muthén, L. K., & Muthén, B. O. (1998–2010). *M*plus *user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA; Sage.

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling, 16*, 583–601.

Skinner, C. J., Holt, D., & Smith, T. M. F. (Eds.). (1989). *Analysis of complex surveys.* New York, NY: Wiley.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Stapleton, L. M. (2006). Using multilevel structural equation modeling technique with complex sample data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 345–383). Greenwich, CT: Information Age.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48,* 817–838.

Yuan, K.-H., & Bentler, P. M. (2002). On normal theory based inference for multilevel models with distributional violations. *Psychometrika, 67*, 539–562.

Yuan, K.-H., & Bentler, P. M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models. *Sociological Methodology, 37*, 53–82.