The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	Applications of multilevel structural equation modeling to cross-cultural research
Author(s)	Cheung, MWL; Au, K
Citation	Structural Equation Modeling, 2005, v. 12 n. 4, p. 598-619
Issued Date	2005
URL	http://hdl.handle.net/10722/53183
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Applications of Multilevel Structural Equation Modeling to Cross-Cultural Research

Mike W.-L. Cheung
The University of Hong Kong

Kevin Au
The Chinese University of Hong Kong

Multilevel structural equation modeling (MSEM) has been proposed as an extension to structural equation modeling for analyzing data with nested structure. We have begun to see a few applications in cross-cultural research in which MSEM fits well as the statistical model. However, given that cross-cultural studies can only afford collecting data from a relatively small number of countries, the appropriateness of MSEM has been questioned. Using the data from the International Social Survey Program (1997; N = 15,244 from 27 countries), we first showed how Muthén's MSEM procedure could be applied to a real data set on cross-cultural research. Given a small country-level sample size (27 countries) we then demonstrated that results on the individual level were quite stable even when using small individual-level sample sizes, whereas the group-level parameter estimates and their standard errors were affected unsystematically by varying individual-level sample sizes. Use of the findings for cross-cultural research and other areas with limited numbers of groups are discussed.

Cross-cultural research has initiated many studies on how and why people behave differently across nations. Culture has been widely used as a variable to explain many cross-national behavioral and psychological differences. Conceptually speaking, multilevel structural equation modeling (MSEM) seems to be the immediate choice of statistical model for a variety of research questions in cross-cultural research. However, MSEM may require large group-level samples of about 100

(Hox & Maas, 2001). This requirement is next to impossible for cross-cultural research, as there are only 200 countries in the world, many of which are small and developing, and it is costly and timely to collect cross-national data. This may be one of the reasons applications of MSEM in cross-cultural research are rare, even though multilevel models have been introduced into cross-cultural research (e.g., Chao, 2000; van de Vijver & Leung, 2000, 2001).

Another possible reason why MSEM in not popular in cross-cultural research is that the model is often seen as complicated and esoteric to many applied-oriented researchers. The statistical techniques of MSEM have been available for more than a decade (e.g., Goldstein & McDonald, 1988; B. O. Muthén, 1989), but its implementation remained quite daunting to researchers. Recent developments in structural equation modeling (SEM) packages, such as LISREL (Jöreskog & Sörbom, 1996), EQS (Bentler, 1995), and M*plus* (L. K. Muthén & Muthén, 2001), have, however, eased some of the more complicated aspects of MSEM.

To shed some light on the applicability of MSEM to cross-cultural research, this study has two objectives. First, MSEM is applied in a typical cross-cultural data set making use of procedures introduced by B. O. Muthén (1989, 1994) that can be applied in nearly all SEM software. By showing that MSEM can now be applied with ease, we hope to demonstrate its usefulness and encourage cross-cultural researchers to use the model. Second, to provide some hints on the sample size issue when applying MSEM on cross-cultural research, the performance of MSEM is investigated by varying different sample sizes of individual-level data. By showing the effects of different sample sizes, we hope to provide some practical suggestions for cross-cultural research as well as other areas in which the number of groups may be limited.

MULTILEVEL STRUCTURAL EQUATION MODELING

Levels Issues in Data Analysis

Human behavior is complicated and multilevel in nature (Klein, Tosi, & Cannella, 1999; Kozlowski & Klein, 2000), prompting considerable debate in many disciplines (see Chan, 1998; House, Rousseau, & Thomas-Hunt, 1995; Klein, Dansereau, & Hall, 1994; Klein & Kozlowski, 2000; Rousseau, 1985). As pointed out by Klein et al. (1999), human behavior is primarily studied within a micro–macro framework, with neither approach fully recognizing the multilevel nature of the data.

Generally, there are three approaches to analyzing data with nested structures: disaggregation, aggregation, and multilevel models (Hofmann, 1997). The first approach is one in which only individual-level data are used in the analysis. In cross-cultural literature, disaggregation factor analysis is termed *pancultural factor analysis,* factor analysis based on only individual-level data disregarding the

country grouping (Triandis et al., 1986). Many researchers (e.g., Hofmann, 1997; Raudenbush & Bryk, 2002) criticized this approach as the random sampling assumption required by most statistical techniques is often violated. The standard errors of parameter estimates are underestimated and the test statistics are inflated; that is, Type I errors are much higher than the nominal values (Raudenbush & Bryk, 2002). Thus, conclusions based on analyzing disaggregated data are usually incorrect.

The second approach is to aggregate data generated from an individual level into the group level. The aggregated data are used in the analysis only. This type of analysis is also known as *ecological analysis.* The most famous example in cross-cultural research is Hofstede's (1980) ecological factor analysis on data from 117,000 samples of 53 nations and regions. Based on the factor analysis on the country means, Hofstede proposed four cultural dimensions, which are frequently used to explain cross-cultural differences in a variety of disciplines. However, this approach has several problems. First, the resulting sample size is much smaller than the original individual-level data. Second, aggregated data may not always be a fair representation of group-level data (Klein & Kozlowski, 2000). Third, information provided from the aggregated means may ignore potentially meaningful individual-level variations (Hofmann, 1997), such as intracultural variation (Au, 1999; Au & Cheung, 2005) or within-group dispersion (Chan, 1998; Klein, Conn, Smith, & Sorra, 2001). Fourth, researchers may commit ecological fallacy, a logical fallacy inherent in making causal inferences from group-level data to individual-level behaviors (Robinson, 1950). Hence, B. O. Muthén (1994) concluded that factor analysis based on aggregated data is often misleading, if not incorrect.

The third approach is multilevel models. These models incorporate individual and group-level data and differentiate group-level and individual-level effects (Raudenbush & Bryk, 2002). Similar to the applications of the hierarchical linear model to regression in the context of the multilevel model, MSEM is a direct generalization of SEM in the context of the multilevel model (e.g., Bentler & Liang, 2003; Goldstein & McDonald, 1988; Lee, 1990; B. O. Muthén, 1989, 1994; Raudenbush, 1993).[1] Applications of MSEM to cross-cultural research are still rare (see, e.g., Dyer, Sipe, & Hanges, 1997).

Muthén's Model

The most general approach to estimating multilevel data is full information maximum likelihood (FIML) because of its asymptotic optimality, meaning its estima-

---

[1]Multilevel models and MSEM were developed in different disciplines with different focuses. Recently, several researchers, such as Bauer (2003), Curran (2003), and Rovine and Molenaar (1998, 2000), among others, have suggested that multilevel models can be integrated within the general framework of SEM.

tors have the smallest possible standard errors among other estimators (Bentler, 2001). However, FIML is computationally demanding to use practically, especially when applied to unbalanced data (i.e., unequal group size). The model specifications with the conventional SEM software can also be very tedious (Bentler, 2001; B. O. Muthén, 1989). To offset this, B. O. Muthén (1989, 1994) proposed a simpler estimator called Muthén's maximum-likelihood-based (MUML) estimator.[2]

The MUML estimator is probably the most widely used procedure for applied researchers conducting MSEM (e.g., Duncan, Alpert, & Duncan, 1998; Dyer et al., 1997; Kaplan, & Elliott, 1997a, 1997b) because it can be easily implemented in nearly all SEM software packages without modifications. Empirical findings suggest that the MUML estimator gives similar results as FIML estimators with rough approximations to the correct chi-square test statistics and standard errors of parameter estimates (Hox, 1993; Hox & Maas, 2001; McDonald, 1994; B. O. Muthén 1989, 1994). Thus, we use the MUML estimator to illustrate the procedures of MSEM in this study.

In analyzing multilevel data, observed scores at the individual level y_{gi} are decomposed into a between-group component y_g, which equals the aggregated group mean, and a within-group component y_w, which equals the centered score from the corresponding group mean. That is,

$$y_{gi} = y_g + y_w \tag{1}$$

Because y_g and y_w are uncorrelated, the total population covariance matrix (Σ_T) can be decomposed into a between-group population covariance matrix (Σ_B) and a pooled within-group population covariance (Σ_W),

$$\Sigma_T = \Sigma_B + \Sigma_W \tag{2}$$

and their sample estimates are S_T, S_B and S_{PW}, where PW = pooled within-group:

$$S_T = (N-1)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{Ng} (y_{gi} - \overline{y})(y_{gi} - \overline{y})' \tag{3}$$

$$S_B = (G-1)^{-1} \sum_{g=1}^{G} N_g (y_g - \overline{y})(y_g - \overline{y})' \tag{4}$$

and

$$S_{PW} = (N-G)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{Ng} (y_{gi} - \bar{y}_g)(y_{gi} - \bar{y}_g)' \qquad (5)$$

where G and N are the total number of group and sample sizes, respectively. B. O. Muthén (1989) showed that S_T and S_{PW} are the consistent and unbiased estimators for Σ_T and Σ_W, and S_B is a consistent and unbiased estimator of

$$\Sigma_W + c\Sigma_B, \qquad (6)$$

where c is a function of the group sizes,

$$c = [N^2 - \sum_{g=1}^{G} N_g^2][N(G-1)]^{-1} \qquad (7)$$

c is the common group size for balanced data. For unbalanced data, it is close to the mean of the group sizes when a large number of groups is used.

S_T, S_B, and S_{PW} can be calculated using standard statistical packages or SEM software.[3] Then S_B and S_{PW} can be modeled as a two-group SEM problem with suitable cross-group constraints. Because S_{PW} is an unbiased estimator of Σ_W, we can propose a within-structure model $\Sigma_W(\theta)$ on S_{PW} directly. As S_B is a weighted sum of Σ_W and Σ_B, a between- and a within-structure model $\Sigma_W(\theta)$ and $\Sigma_B(\theta)$ are required at the same time to model S_B properly. Because the within-structure model is the same in S_B and S_{PW}, appropriate cross-group constraints are required. B. O. Muthén (1989) suggested a four-step procedure before analyzing MSEM.

Step 1: Conventional structural equation model of the total sample covariance matrix.   In this analysis, the multilevel nested data structure is ignored and conventional structural equation model is fitted. The purpose of this step is to detect obvious misspecifications of the proposed model. However, the goodness of fit of the model should not be interpreted on its own, especially for data with large intraclass correlations, large class sizes, or highly correlated variables.

Step 2: Estimation of between-group variation.   If the between-group variations are not present (i.e., $\Sigma_B = 0$), MSEM can be simplified into conventional SEM. However, as suggested by B. O. Muthén (1994), an easier way to check if $\Sigma_B = 0$ is to estimate the intraclass correlation or ICC(1) for each item, which is defined as

$$\frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \qquad (8)$$

---

[3]EQS 6, LISREL 8.54, and M*plus* 1 (or later versions) can create the between- and within-group covariance matrices easily. Other statistical packages, such as SPSS or SAS, can also be used to calculate the within- and between-group covariance matrices.

where σ_B^2 and σ_W^2 are the between- and within-group variances, respectively. If ICCs(1) for the variables are not large, conventional SEM analyses are generally reasonable and unbiased (Julian, 2001).

Step 3: Estimation of pooled within-group structure. S_{PW} is used to estimate the pooled within-group structure because it is an unbiased estimator of Σ_W. Then the proposed factor structure at the individual level, $\Sigma_W(\theta)$, can be proposed to fit the S_{PW}.

Step 4: Estimation of between-group structure.   As pointed out by Chan (1998), Klein and Kozlowski (2000), and B. O. Muthén (1994), among others, the aggregated variables may not have the same meaning as those in the individual-level data; hence, the between-structure model may not be the same as that of the within-structure model. From Equation 6, the MUML estimate of Σ_B is

$$c^{-1}(S_B^2 - S_{PW}^2). \tag{9}$$

The MUML estimate of Σ_B is frequently not positively definite, which is usually required by most estimation methods. B. O. Muthén's (1994) experience suggested that the analysis of S_B or the estimate of Σ_B based on Equation 9 gives similar results. Thus, S_B can be used as a convenient substitute for the MUML estimate in most cases.

Analysis of multilevel structural equation model.   Based on the results from the previous steps, MSEM can be proposed and analyzed with a two-group SEM approach. By analyzing S_B and S_{PW} simultaneously, model fit and parameter estimates can be assessed. If the proposed model does not fit the data, standard model modification steps can be used to search for a more meaningful and better fitted model.

Sample Size Issues in MSEM

When applying a sophisticated statistical method, such as MSEM, practical issues are usually of concern. A fundamental issue is the minimum sample sizes required at the individual and group levels. Because MSEM is quite new, only a few studies have investigated its sample size requirements. Hox and Maas (2001) assessed the robustness of the MUML estimator in unequal groups and small sample sizes at both individual and group levels with a simulation study. They found that problems would occur for small, inadequate group-level samples. They suggested that the

group-level sample size should be at least 100 for good performance of the MUML estimator.

Illustrated examples using MSEM in the literature vary a lot in the sample sizes employed. For instance, B. O. Muthén (1994) used a data set with 3,724 students nested within 197 classes. Duncan et al. (1998) analyzed 1,076 adolescents from 450 households, and Kaplan and Elliott (1997a, 1997b) analyzed 1,165 students from 356 schools. These studies satisfy the suggestion by Hox and Maas (2001). However, there are still some studies that do not fulfill the suggestion of group-level sample size larger than 100. For instance, Heck (2001) used 384 participants from 56 groups, whereas Hox (1993) used only 187 children from 37 families.

The issue of small group-level sample size becomes even more critical when applying MSEM to cross-cultural research because the cultural-level sample size is always small, whereas the individual-level sample size is comparatively large. For instance, the exploratory analyses on comparing the country and individual factor structures conducted by van de Vijver and Poortinga (2002) involved 39 regions and 47,871 respondents, and van Hemert, van ve Vijver, Poortinga, and Georgas's (2002) study used a data set with 25,922 respondents from 24 countries. The MSEM study of Dyer et al. (1997) consisted of 1,817 respondents from 17 countries. It is easy to observe that the typical group-level sample size is around 20 to 40, whereas the individual-level sample size can be as large as 100 to 1,000 per country in cross-cultural research.

Although the group-level sample size is quite small in cross-cultural research, it is unclear whether the large individual-level sample size is beneficial to the overall performance of MSEM, especially the performance at the group level. Regarding multilevel regression, some researchers suggest a trade-off between sample sizes at different levels (Cohen, 1998; Mok, 1995; Raudenbush & Liu, 2000; Snijders & Bosker, 1993). That is, increasing individual-level sample size may reduce the group-level sample size requirement. The idea is that increasing individual-level sample size may increase the precision on the estimates of the group means. Thus, S_B will be closer to its population value given the same number of group-level samples. Although most findings suggest that increasing the group-level sample size is more beneficial than increasing the individual sample size (e.g., Hox & Maas, 2001; Snijders & Bosker, 1993), it is not clear whether the large individual-level sample size in cross-cultural research will help the group-level model fit or parameter estimates in MSEM. Moreover, most of these suggestions on minimum sample sizes (e.g., Hox & Maas, 2001) were based on simulation studies with artificial data. The effects of sample sizes on real data are hardly known. To shed light on this issue, we selected several subsamples of individuals that are different in size from a real data set and tested the performance of the MUML estimator.
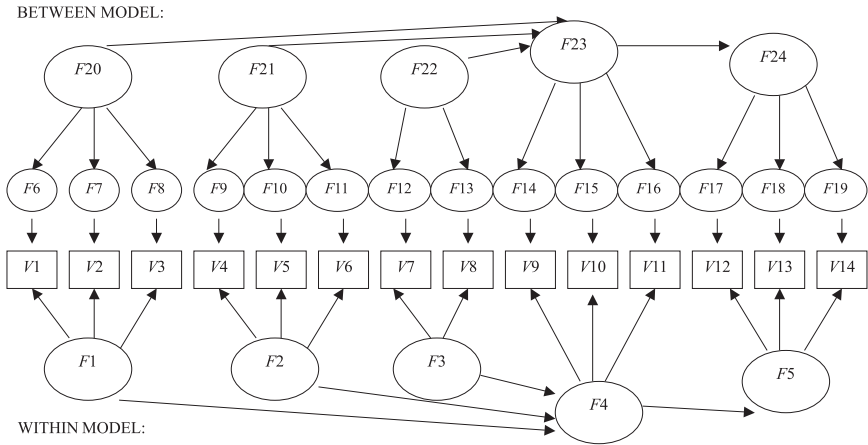
FIGURE 1   A multilevel structural equation model on intention to quit with organizational commitment as the mediator.
Note. $F1$ to $F5$ represent job prospects, job nature, citizenship behaviors, organizational commitment, and intention to quit, respectively. The loadings from latent factors ($F6$–$F19$) to their corresponding observed variables ($V1$–$V14$) are fixed at the square root of the scaling parameter ($\sqrt{c}$) or 23.656 in the between-group model. For simplicity, the measurement errors, disturbances, and covariances among exogenous latent variables are not shown in the figure.

ILLUSTRATION WITH REAL DATA

Data and Sample

A data set from the International Social Survey Program (ISSP, 1997) was used to demonstrate the use of MSEM on cross-cultural research. The complete sample without missing data included 15,244 full-time employees from 27 countries. The sample size per country is shown in Table 1. Data collection in most countries used probability-based samples of adults, and a few countries used quota sampling. The questionnaires were specially designed to measure work-oriented variables, such as importance and perception of work attitudes. Thus, it is most suitable for comparing work-related attitudes across countries. Other researchers have successfully used data sets from ISSP for a variety of purposes (e.g., Blanchflower & Freeman, 1997; Cheung & Rensvold, 1999), suggesting that the data are of good quality. This data set is also typical of those generated from other large-scale studies (e.g., Abramson & Inglehart, 1995; House, Javidan, Hanges, & Dorfman, 2002). If anything, its sample sizes, at both the individual and group level, tend to be small in comparison to other studies. Implications derived from this example are likely to be generalizable to those with similar sample sizes in cross-cultural research.

As shown in Figure 1, we conceptualized a five-factor model on job attitudes based on the work attitude literature (e.g., Karasek, 1979). These factors were measured by 14 items:

Job prospects ($F1$)—job security ($v1$), high income ($v2$), advancement ($v3$).
Job nature ($F2$)—interesting job ($v4$), independent work ($v5$), flexible working hours ($v6$).
Citizenship behaviors ($F3$)—job helps people ($v7$), useful to society ($v8$).
Organizational commitment ($F4$)—work harder to help the firm ($v9$), proud working with the firm ($v10$), stay with the firm even with higher pay by other firms ($v11$).
Intention to quit ($F5$)—days absent in last 6 months ($v12$), likely to find another job ($v13$), worry of losing job ($v14$).

Three factors, namely job prospects ($F1$), job nature ($F2$), and citizenship behaviors ($F3$), were proposed to influence intention to quit ($F5$) indirectly through organizational commitment ($F4$), with organizational commitment acting as the mediator. Detailed item descriptions can be found in ISSP (1997). Instead of using a confirmatory factor analytic model as an illustration, we used a full structural equation model because a model with structural relations is more useful in cross-cultural research.

The individual-level structure model was transformed into the country-level structure model by the direct consensus model, which "uses within-group consensus of the lower level units as the functional relationship to specify how the construct conceptualized and operationalized at the lower level is functionally isomorphic to another form of the construct at the higher level" (Chan, 1998, p. 237; see also Klein et al., 2001). In direct consensus models, high internal agreement is required and consistent with the need of high intraclass correlation in MSEM.

To study the effects of sample sizes on MSEM, we randomly selected subsamples without replacement with different sample sizes (50, 100, 200, 300, 400, and 500) from each country. Then, we fitted the model already proposed. In this resampling procedure, the number of countries was fixed but the individual-level sample varied. This is quite similar to the actual data collection in cross-cultural research in which it is much more difficult to collect country-level data than it is to recruit individual participants within a particular country. Because the sample sizes are unbalanced (see Table 1), some countries do not have enough data for large samples (e.g., $N = 400$). This unbalanced sample size situation is also realistic in cross-cultural research in which the samples collected are usually unbalanced across countries.

TABLE 1
Samples Sizes in Individual Country or Region

Countries/Regions	Sample Sizes
Germany East	196
Israel (Arabs)	227
Bangladesh	305
Poland	325
Spain	344
Bulgaria	389
Italy	415
Philippines	415
Japan	422
Great Britain	460
Slovenia	460
Cyprus	467
Czech Republic	473
France	509
Germany West	517
Canada	535
Russia	542
Israel (Jews)	557
Hungary	580
Sweden	634
Denmark	643
New Zealand	659
United States	708
Portugal	846
Netherlands	1,013
Norway	1,262
Switzerland	1,341

Results: Application of MSEM With the Four-Step Data Preparation Process

Step 1: Conventional structural equation model of the total sample covariance matrix.   First, we fitted the model ignoring the nested structure of the data with EQS 6 (Bentler, 2001). The proposed model fit the data inadequately, $\chi^2(70) = 3,474$, $p < .001$, $N = 15, 217$, the comparative fit index (CFI) = 0.89, Bentler–Bonett Normed Fit Index (NFI) = 0.89, and root mean squared error of approximation (RMSEA) = 0.057. By ignoring the nested structure of the data, we know little from the rejection of the proposed model. It may be due to the nested structure of the data or because the proposed model does not fit the data adequately, or even both reasons. Moreover, as suggested by B. O. Muthén (1994) these results should not be interpreted on their own.

TABLE 2
Total Sample Covariance Matrix and Intraclass Correlations of the Variables

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
V1	0.514													
V2	0.169	0.681												
V3	0.166	0.284	0.887											
V4	0.039	0.063	0.180	0.444										
V5	0.024	0.061	0.168	0.210	0.773									
V6	–0.004	0.133	0.104	0.107	0.244	1.099								
V7	0.098	0.065	0.160	0.144	0.191	0.143	0.720							
V8	0.111	0.075	0.168	0.113	0.134	0.141	0.464	0.762						
V9	0.022	–0.020	0.104	0.054	0.101	–0.023	0.084	0.102	1.091					
V10	0.036	–0.021	0.104	0.069	0.073	–0.009	0.138	0.151	0.503	0.893				
V11	–0.003	–0.070	–0.049	0.012	0.034	–0.016	0.099	0.105	0.345	0.436	1.497			
V12	–0.009	0.020	–0.016	–0.026	–0.046	0.034	0.017	0.004	–0.100	–0.080	–0.074	1.223		
V13	–0.023	0.047	0.069	0.017	–0.012	0.055	–0.026	–0.030	–0.167	–0.274	–0.347	0.093	0.990	
V14	0.102	0.123	0.058	–0.036	–0.050	0.020	0.028	0.048	–0.062	–0.090	–0.111	0.043	0.195	1.009
Intraclass correlations	0.067	0.219	0.176	0.067	0.084	0.051	0.058	0.076	0.077	0.068	0.039	0.024	0.046	0.138

Note. N = 15,217. All intraclass correlations are significant at $p < .001$.

Step 2: Estimation of between-group variation. Table 2 shows the total covariance matrix and the ICC(1) of the variables. The ICC(1) of the variables ranged from .02 (days absent in last 6 months, $v12$) to .22 (high income, $v2$). These indicate that the country variation can account for 2% to 22% of the variance of the variables. After reviewing a number of studies, James (1982) found that ICC(1) values ranged from .00 to .50 with a median value of .12. Bliese (2000) criticized James's approach as overestimating both the range and median of ICC(1) because eta-squared was equated with ICC(1). Bliese (2000) put the typical values at between .05 and .20 based on his experience.

On the other hand, B. O. Muthén's (1994) experience with survey data suggested that common values of ICC(1) ranged from .00 to .50. The differences among them may be due to the difference of their disciplines. Moreover, the ICC(1) may underestimate the true intraclass correlation because of the individual-level measurement error (B. O. Muthén, 1994). Our ICC(1) results (.02–.22) are quite typical of others' findings. The results indicate that the data are not independent. MSEM is necessary for making valid statistical inferences.

Step 3: Fitting pooled-within structure model. By using EQS 6 (Bentler, 2001) the total covariance matrix was partitioned into pooled within- and between-sample covariance matrices (see Table 3). The square root of the ad hoc estimator constant or the scaling parameter (c in Equation 7) was 23.66. By analyzing S_W, the proposed model fitted the data marginally well, $\chi^2(70) = 2,820$, $p < .001$, $N = 15, 217$, CFI $= 0.91$, NFI $= 0.90$, RMSEA $= 0.051$. These results are consistent with the general findings that analyzing S_W gives better results than analyzing S_T.

Step 4: Fitting between structure model. As the sample of group level is not large ($G = 27$), analyzing S_B alone may not be adequate. Hence, this analysis is integrated into the MSEM.

Step 5: Fitting multilevel structural equation model. The MSEM fitted the data reasonably well with $\chi^2(140) = 2,956$, $p < .001$, $N = 15, 217$, CFI $= 0.90$, NFI $= 0.90$, and especially RMSEA $= 0.036$. These results show that the proposed model fits better under a multilevel framework. The parameter estimates and standard errors of parameter estimates are shown in Table 4.

Several observations are made. When comparing MSEM with SEM ignoring the nested structure, the reduction of chi-square is large ($\Delta\chi^2 = 518$) when compared to the loss of degrees of freedom ($\Delta df = 70$). Although the formal chi-square difference test is not appropriate in testing their significance, it is not difficult to see that MSEM fits better than SEM ignoring the nested structure by checking their chi-squares and other goodness-of-fit indexes.

TABLE 3
Pooled Within-Sample and Between-Sample Correlation Matrices of the Variables

	V1	*V2*	*V3*	*V4*	*V5*	*V6*	*V7*	*V8*	*V9*	*V10*	*V11*	*V12*	*V13*	*V14*
V1		0.530	0.602	–0.053	–0.052	0.031	0.242	0.431	0.113	0.196	0.189	–0.291	0.030	0.503
V2	0.260		0.448	–0.182	–0.192	0.518	0.379	0.547	–0.110	–0.072	0.101	0.125	0.208	0.602
V3	0.208	0.346		0.132	0.212	0.130	0.638	0.639	0.303	0.444	0.108	–0.150	0.232	0.244
V4	0.092	0.157	0.310		0.608	0.274	0.055	–0.315	–0.161	0.147	–0.201	–0.305	0.233	–0.182
V5	0.044	0.128	0.203	0.339		0.171	0.197	–0.128	0.172	0.147	–0.137	–0.276	–0.077	–0.118
V6	–0.007	0.116	0.105	0.146	0.272		0.151	0.101	–0.288	–0.220	–0.078	0.185	0.224	0.271
V7	0.156	0.060	0.156	0.269	0.261	0.161		0.823	0.406	0.687	0.334	0.212	–0.114	0.164
V8	0.159	0.042	0.151	0.232	0.201	0.158	0.613		0.377	0.581	0.409	0.141	–0.006	0.342
V9	0.023	–0.012	0.082	0.095	0.104	–0.003	0.073	0.090		0.601	0.477	0.198	–0.226	–0.206
V10	0.043	–0.022	0.079	0.106	0.084	0.004	0.138	0.153	0.503		0.556	0.017	–0.131	–0.128
V11	–0.014	–0.090	–0.057	0.027	0.042	–0.009	0.084	0.081	0.259	0.369		0.266	–0.051	0.127
V12	0.001	0.015	–0.006	–0.025	–0.037	0.023	0.010	–0.003	–0.101	–0.081	–0.065		–0.208	–0.207
V13	–0.036	0.042	0.059	0.014	–0.010	0.044	–0.026	–0.037	–0.157	–0.302	–0.295	0.095		0.162
V14	0.106	0.057	0.029	–0.040	–0.050	–0.003	0.021	0.023	–0.043	–0.092	–0.109	0.055	0.200	
WS	0.693	0.734	0.859	0.645	0.843	1.022	0.825	0.841	1.006	0.914	1.201	1.093	0.97	0.936
BS	4.456	9.164	9.378	4.117	6.086	5.711	4.919	5.747	6.921	5.911	5.813	4.208	5.164	8.886

Note. The lower and upper matrices are the pooled within-sample and between-sample correlation matrices, respectively. WS and BS stand for the pooled within-sample and between-sample standard deviations, respectively.

TABLE 4
Parameter Estimates and Their Standard Errors of MSEM

	MSEM (Within)	MSEM (Between)
Factor loadings		
$V1, F1$	0.243 (0.007)	0.120 (0.035)
$V2, F1$	0.385 (0.008)	0.270 (0.071)
$V3, F1$	0.581 (0.010)	0.316 (0.070)
$V4, F2$	0.386 (0.007)	0.171 (0.024)
$V5, F2$	0.485 (0.009)	0.155 (0.046)
$V6, F2$	0.349 (0.010)	0.063[b] (0.047)
$V7, F3$	0.688 (0.009)	0.169 (0.033)
$V8, F3$	0.619 (0.009)	0.240 (0.034)
$V9, F4$	1.000[a]	1.000[a]
$V10, F4$	1.286 (0.027)	1.420 (0.379)
$V11, F4$	0.937 (0.021)	0.780 (0.312)
$V12, F5$	1.000[a]	1.000[a]
$V13, F5$	4.220 (0.352)	−1.150[b] (1.413)
$V14, F5$	1.522 (0.121)	−1.933[b] (2.364)
Structural paths		
$F4, F1$	−0.033 (0.010)	−0.118[b] (0.076)
$F4, F2$	0.073 (0.013)	0.098 (0.045)
$F4, F3$	0.090 (0.009)	0.226 (0.096)
$F5, F4$	−0.142 (0.012)	0.091[b] (0.158)
Factor correlations		
$F2, F1$	.554 (0.013)	−.003[b] (0.226)
$F3, F1$	.270 (0.012)	.776 (0.114)
$F3, F2$	.534 (0.011)	−.325[b] (0.180)

Note. $N = 15,217$.
[a]They are fixed as 1.00 for identification purpose. [b]They are not statistically significant at $p < .05$; all other parameter estimates are significant at $p < .05$.

Second, the parameter estimates for the pooled within sample and between sample are not necessarily the same. For instance, $v13$ (likely to find another job) and $v14$ (worry of losing job) are good indicators for the intention to quit at the individual level, but they are not good indicators at the country level, as they are not statistically significant. Indeed, the construct of intention to quit is problematic at the country level as none of its indicators are significant. This provides some indication to further investigate the applicability of intention to quit at the country level. The negative relation between organizational commitment and intention to quit at the individual level suggests that people with higher organizational commitment are less likely to quit their jobs. However, this relation does not hold at the country level. On the other hand, the relation between job prospects and citizenship behaviors is stronger at the country level than at the individual level. There-

TABLE 5
Goodness-of-Fit Indexes of the Resampling Results

Sample Size per Country	Total Sample Size	χ^2	CFI	NFI	RMSEA
50	1,350	393	0.91	0.87	0.037
100	2,700	679	0.90	0.89	0.038
200	5,396	1,078	0.91	0.90	0.036
300	7,827	1,546	0.91	0.90	0.036
400	10,186	1,932	0.91	0.91	0.035
500	11,898	2,206	0.91	0.91	0.035

Note.   The degrees of freedom of the models are 140 and all the χ^2 are significant at $p < .001$. CFI = Comparative Fit Index; NFI = Normed Fit Index; RMSEA = Root Mean Squared Error of Approximation.

fore, SEM ignoring nested nature may give incorrect substantive interpretations concerning the relations among these constructs.

Results: Effects of Sample Size

The goodness-of-fit indexes of the models with different sample sizes are shown in Table 5. First, as expected, the χ^2 test statistics are directly related to sample size. The differences between small individual-level sample size ($N = 50$ per country) and large individual-level sample size ($N = 500$ per country) in terms of goodness-of-fit indexes are not large. What is interesting is that the sample size per country can be as small as 50, yet the results are still comparable with other large sample size conditions.

Second, the patterns of the goodness-of-fit indexes across group sizes suggest that MSEM is better suited when the individual-level sample size increases. One speculation is that the model fit depends on the between and within models. When the individual-level sample size increases, the within model becomes more stable and fits better. Thus, the overall model fit also increases.

Third, by comparing the goodness-of-fit indexes across different sample sizes per country, they suggest that CFI and RMSEA are insensitive to the sample sizes used, whereas NFI is relatively insensitive to the sample size. However, before concluding that CFI, RMSEA, and NFI can be routinely used in MSEM with small group sizes, more empirical studies are required. Moreover, there are one or two inadmissible estimates (negative variance of the error terms) in the between-structure model, whereas there is no inadmissible estimate in the within-structure model in all sample sizes. The inadmissible estimates in the between-structure model are likely due to the small group sizes (Hox & Maas, 2001). It seems that increasing individual-level sample size may not help much in preventing inadmissible estimates at the group level.

TABLE 6
Parameter Estimates (and Their Standard Errors) of the Pooled-Within Model

Sample Size Per Group	50	100	200	300	400	500
Factor loadings						
$V1, F1$	0.220	0.247	0.220	0.251	0.240	0.239
	(0.024)	(0.017)	(0.012)	(0.010)	(0.009)	(0.008)
$V2, F1$	0.321	0.366	0.360	0.372	0.366	0.369
	(0.025)	(0.018)	(0.013)	(0.011)	(0.010)	(0.009)
$V3, F1$	0.627	0.580	0.582	0.580	0.583	0.591
	(0.038)	(0.024)	(0.018)	(0.014)	(0.013)	(0.012)
$V4, F2$	0.410	0.398	0.422	0.410	0.409	0.398
	(0.022)	(0.016)	(0.011)	(0.009)	(0.008)	(0.007)
$V5, F2$	0.528	0.500	0.514	0.505	0.501(	0.496
	(0.029)	(0.021)	(0.014)	(0.012)	0.010)	(0.010)
$V6, F2$	0.366	0.393	0.375	0.383	0.363	0.365
	(0.034)	(0.024)	(0.017)	(0.014)	(0.012)	(0.011)
$V7, F3$	0.716	0.658	0.722	0.689	0.685	0.684
	(0.028)	(0.020)	(0.015)	(0.012)	(0.010)	(0.010)
$V8, F3$	0.607	0.597	0.603	0.594	0.602	0.617
	(0.027)	(0.020)	(0.014)	(0.011)	(0.010)	(0.009)
$V9, F4$	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
$V10, F4$	1.376	1.204	1.317	1.307	1.287	1.278
	(0.103)	(0.058)	(0.048)	(0.038)	(0.032)	(0.030)
$V11, F4$	0.769	0.796	0.905	0.896	0.884	0.892
	(0.066)	(0.044)	(0.035)	(0.028)	(0.024)	(0.023)
$V12, F5$	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
$V13, F5$	7.705	3.637	4.482	4.187	4.237	4.279
	(3.181)	(0.603)	(0.675)	(0.512)	(0.447)	(0.424)
$V14, F5$	2.445	1.863	1.570	1.571	1.642	1.687
	(0.919)	(0.293)	(0.210)	(0.182)	(0.164)	(0.159)
Structural paths						
$F4, F1$	–0.035[b]	0.008[b]	–0.065	–0.049	–0.047	–0.038
	(0.034)	(0.028)	(0.019)	(0.015)	(0.014)	(0.012)
$F4, F2$	0.029[b]	0.003[b]	0.125	0.077	0.081	0.071
	(0.040)	(0.034)	(0.023)	(0.019)	(0.017)	(0.016)
$F4, F3$	0.133	0.120	0.074	0.104	0.101	0.105
	(0.029)	(0.024)	(0.015)	(0.013)	(0.012)	(0.011)
$F5, F4$	–0.073	–0.142	–0.121	–0.130	–0.128	–0.128
	(0.030)	(0.024)	(0.018)	(0.016)	(0.014)	(0.013)
Factor correlations						
$F2, F1$	0.577	0.584	0.601	0.584	0.597	0.577
	(0.043)	(0.031)	(0.022)	(0.018)	(0.016)	(0.015)
$F3, F1$	0.339	0.345	0.294	0.311	0.305	0.300
	(0.038)	(0.028)	(0.020)	(0.016)	(0.014)	(0.013)
$F3, F2$	0.538	0.561	0.543	0.567	0.566	0.559
	(0.035)	(0.026)	(0.018)	(0.015)	(0.013)	(0.012)

[a]They are fixed as 1.00 for identification purposes. [b]They are not statistically significant at $p < .05$; all other parameter estimates are significant.

TABLE 7
Parameter Estimates (and Their Standard Errors) of the Between Model

Sample Size per Group	50	100	200	300	400	500
Factor loadings						
V1, F1	0.037	0.054	0.075	0.096	0.101	0.104
	(0.014)	(0.016)	(0.022)	(0.026)	(0.028)	(0.029)
V2, F1	0.045[b]	0.092	0.110	0.128	0.132	0.146
	(0.026)	(0.033)	(0.046)	(0.055)	(0.061)	(0.065)
V3, F1	0.126	0.142	0.229	0.287	0.336	0.362
	(0.025)	(0.033)	(0.041)	(0.050)	(0.056)	(0.051)
V4, F2	0.049	0.078	0.111	0.139	0.155	0.125
	(0.009)	(0.012)	(0.026)	(0.020)	(0.022)	(0.041)
V5, F2	0.028[b]	0.067	0.092	0.115	0.136	0.181
	(0.015)	(0.019)	(0.034)	(0.035)	(0.041)	(0.060)
V6, F2	0.015	0.025[b]	0.050[b]	0.059	0.059	0.072
	(0.021)	(0.023)	(0.037)	(0.038)	(0.043)	(0.051)
V7, F3	0.043	0.073	0.102	0.108	0.132	0.172
	(0.011)	(0.016)	(0.022)	(0.025)	(0.029)	(0.029)
V8, F3	0.075	0.113	0.151	0.171	0.197	0.183
	(0.013)	(0.017)	(0.022)	(0.025)	(0.030)	(0.036)
V9, F4	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
V10, F4	0.898	1.010	0.944	0.717	1.227	1.040
	(0.352)	(0.372)	(0.269)	(0.215)	(0.322)	(0.280)
V11, F4	1.085	0.802	0.818	0.662	0.661	0.674
	(0.414)	(0.345)	(0.273)	(0.229)	(0.287)	(0.245)
V12, F5	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]	1.000[a]
V13, F5	−0.977[b]	−0.039[b]	−0.348[b]	−1.008[b]	−1.137[b]	−1.171[b]
	(1.037)	(0.292)	(0.625)	(0.853)	(1.082)	(1.161)
V14, F5	−5.050[b]	−0.346[b]	−1.295[b]	−1.561[b]	−2.553[b]	−1.823[b]
	(8.897)	(0.497)	(1.701)	(1.328)	(2.565)	(1.804)
Structural paths						
F4, F1	0.009[b]	−0.034[b]	−0.031[b]	−0.034[b]	−0.027[b]	−0.021[b]
	(0.031)	(0.031)	(0.052)	(0.064)	(0.048)	(0.050)
F4, F2	0.014[b]	0.016[b]	0.040[b]	0.028[b]	0.069[b]	−0.006[b]
	(0.026)	(0.021)	(0.039)	(0.048)	(0.040)	(0.040)
F4, F3	0.030	0.080	0.127	0.144	0.138	0.169
	(0.038)	(0.037)	(0.062)	(0.073)	(0.065)	(0.066)
F5, F4	0.038	0.550[b]	0.220[b]	0.176[b]	0.112[b]	0.123
	(0.119)	(0.290)	(0.211)	(0.149)	(0.147)	(0.145)
Factor correlations						
F2, F1	0.062[b]	0.185[b]	0.152[b]	0.113[b]	0.176[b]	0.350[b]
	(0.234)	(0.233)	(0.216)	(0.206)	(0.196)	(0.207)
F3, F1	0.707	0.614	0.702	0.672	0.657	0.680
	(0.153)	(0.166)	(0.127)	(0.129)	(0.133)	(0.120)
F3, F2	−0.454	−0.218[b]	−0.327[b]	−0.369	−0.306	0.213[b]
	(0.217)	(0.210)	(0.201)	(0.178)	(0.186)[b]	(0.234)

[a]They are fixed as 1.00 for identification purposes. [b]They are not statistically significant at $p < .05$; all other parameter estimates are significant.

The parameter estimates and their standard errors for the pooled-within and between models are shown in Tables 6 and 7. When comparing the parameter estimates and their standard errors for the pooled-within model in Table 6, several results are observed. First, the parameter estimates of the pooled-within analysis are quite comparable with each other even when the sample size is small ($N = 50$ per country). There are only a few discrepancies across the sample sizes. For instance, the structural paths from job security ($F1$) to organizational commitment ($F4$) and job nature ($F2$) to organizational commitment ($F4$) are nonsignificant at $N = 50$ and $N = 100$ per group but are all significant at $N \geq 200$. This indicates that large individual-level sample size is still preferable even though other goodness-of-fit indexes are similar. The standard errors of the parameter estimates are also consistently smaller when the sample sizes get larger.

For the parameter estimates and their standard errors in the between-group analysis in Table 7, no consistent pattern emerges when increasing the individual-level sample sizes. Some parameter estimates, for instance, ($v1$, $F1$) and ($v2$, $F1$), turn out to be larger when the sample size per group increases, whereas other parameter estimates, for instance, ($v11$, $F4$), become smaller when the sample size per group increases. The increase of Level 1 sample sizes can have different effects on the parameter estimates and standard errors on Level 2.

SUMMARY AND DISCUSSION

By using the data from ISSP (1997), we showed how B. O. Muthén's (1989, 1994) procedure could be applied in EQS for analyzing multilevel data. We demonstrated that MSEM can be used to analyze a cross-cultural data set when the group-level sample size is relatively small. It may also be useful in such a situation for other kinds of groups, such as school districts and training programs. Further analysis resulted in an observation concerning application. With respect to model fit, parameter estimates, and their standard errors at the individual level, the required sample sizes for individual level need not be very large. The model fit of sample sizes of 100 or even 50 per country is quite comparable with 500 or even more per country. Consistent with statistical theory, of course, larger sample sizes still result in smaller standard errors at the individual level.

One important finding is that using different individual-level sample sizes has a nonsystematic impact on the parameter estimates and their standard errors at the group level. In other words, increasing the individual-level sample size does not necessarily benefit the parameter estimates and their standard errors at the group level. Hence, researchers should not increase sample sizes at the individual level and hope that they can help the estimates at the group level. The current findings indicate that statistical inferences at the individual level for MSEM are quite good even with a small number of countries. However, it is not clear whether statistical

inferences at the group level can be drawn properly in cross-cultural research when a small number of countries are used.

Although it may not be possible to increase the number of countries in cross-cultural research, one possible resolution might be to increase the group-level sample size by using regions instead of countries as the sampling unit. As suggested by Au (1999; Au & Cheung, 2005) and others, people differ considerably within one country and this difference varies from city to city (or region to region) within the same country. Thus, it may be more theoretically and practically appealing to define cities (or regions) as the group level unit for cross-cultural research.

Before concluding, one limitation of this study should be addressed. Because the data set was based on a large-scale study collaborated on by researchers in different countries, the resultant individual-level sample sizes were varied. As suggested by Hox and Maas (2001), unbalanced sample sizes may also influence the performance of MSEM. Further studies are required to clarify the unbalanced sample size issues in cross-cultural research.

All in all, MSEM is a powerful tool for the study of cross-cultural research. It provides superior estimation to other models that ignore the nested data nature, and it requires only simple programming in common SEM packages. Even though the group-level sample size issue needs further investigation, MSEM is still a viable tool that will hopefully become more popular for testing complex multilevel models and theories in cross-cultural research.

ACKNOWLEDGMENTS

REFERENCES

Abramson, P., & Inglehart, R. (1995). *Value change in global perspective.* Ann Arbor: University of Michigan Press.

Au, K. (1999). Intra-cultural variation: Evidence and implications for international business. *Journal of International Business Studies, 30,* 799–812.

Au, K., & Cheung, M. W. L. (2005). Intra-cultural variation and job autonomy in 42 countries. *Organization Studies, 25,* 1339–1362.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28,* 135–167.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M. (2001). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Liang, J. (2003). Two-level mean and covariance structures: Maximum likelihood via an EM algorithm. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53–70). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Blanchflower, D. G., & Freeman, R. B. (1997). The attitudinal legacy of communist labor relations. *Industrial and Labor Relations Review, 50,* 438–459.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco: Jossey-Bass.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83,* 234–246.

Chao, G. T. (2000). Multilevel issues and culture: An integrative view. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 308–346). San Francisco: Jossey-Bass.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25,* 1–27.

Cohen, M. P. (1998). Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics, 14,* 267–275.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38,* 529–569.

Duncan, T. E., Alpert, A., & Duncan, S. C. (1998). Multilevel covariance structure analysis of sibling antisocial behavior. *Structural Equation Modeling, 5,* 211–228.

du Toit, M., & du Toit, S. H. C. (2001). *Interactive LISREL user's guide.* Lincolnwood, IL: Scientific Software International.

Dyer, N., Sipe, W. P., & Hanges, P. J. (1997). Multilevel confirmatory factor analysis: Demonstration of Muthen's technique. In *Academy of Management best paper proceedings* (pp. 391–394). Madison, WI: Ominpress.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika, 53,* 455–467.

Heck, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New development and techniques in structural equation modeling* (pp. 89–127). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23,* 723–744.

Hofstede, G. (1980). *Culture's consequences.* Beverly Hills, CA: Sage.

House, R., Javidan, M., Hanges, P., & Dorfman, P. (2002). Understanding cultures and implicit leadership theories across the globe: An introduction to project GLOBE. *Journal of World Business, 37,* 3–10.

House, R., Rousseau, D. M., & Thomas-Hunt, M. (1995). The meso paradigm: A framework for the integration of micro and macro organizational behavior. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 17, pp. 71–114). Greenwich, CT: JAI.

Hox, J. J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In J. H. L. Oud & R. A. W. van Blokland-Vogelesang (Eds.), *Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp. 141–156). Nijmegen, Netherlands: ITS.

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.

International Social Survey Program. (1997). *International Social Science Program: Work orientations II, 1997* [Computer file]. Koln, Germany: Zentralarchiv fuer Empirische Sozialforschung.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67,* 219–229.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: A user's reference guide.* Chicago: Scientific Software International.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling, 8,* 325–352.

Kaplan, D., & Elliott, P. R. (1997a). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling, 4,* 1–24.

Kaplan, D., & Elliott, P. R. (1997b). A model-based approach to validating education indicators using multilevel structural equation modeling. *Journal of Educational and Behavioral Statistics, 22,* 323–347.

Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly, 24,* 285–307.

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,* 3–16.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review, 19,* 195–229.

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

Klein, K. J., Tosi, H., & Cannella, A. A., Jr. (1999). Multilevel theory building: Benefits, barriers, and new developments. *Academy of Management Review, 24,* 243–248.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent properties. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco: Jossey-Bass.

Lee, S. Y. (1990). Multilevel analysis of structural equation models. *Biometrika, 77,* 763–772.

McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research, 22,* 399–413.

Mok, M. (1995). *Sample size requirements for 2-level designs in educational research.* Unpublished manuscript, Multilevel Models Project, Institute of Education, University of London, England.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54,* 557–585.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22,* 376–398.

Muthén, L. K., & Muthén, B. O. (2001). *M*plus *user's guide* (2nd ed.). Los Angeles: Muthén & Muthén.

Raudenbush, S. W. (1993). Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical and Statistical Psychology, 48,* 359–370.

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

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods, 5,* 199–213.

Robinson, W. S. (1950). Ecological correlations and the behaviors of individuals. *American Sociological Review, 15,* 351–357.

Rousseau, D. M. (1985). Issues of level in organizational research: Multilevel and cross-level perspectives. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 1–37). Greenwich, CT: JAI.

Rovine, M. J., & Molenaar, P. C. M. (1998). A LISREL model for the analysis of repeated measures with a patterned covariance matrix. *Structural Equation Modeling, 5,* 318–343.

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research, 35,* 51–88.

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18,* 237–259.

Triandis, H. C., Bontempo, R., Betancourt, H., Bond, M. H., Leung, K., Brenes, A., et al. (1986). The measurement of the etic aspects of individualism and collectivism across cultures. *Australian Journal of Psychology, 38,* 257–267.

van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross-Cultural Psychology, 31,* 33–35.

van de Vijver, F. J. R., & Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality, 69,* 1007–1031.

van de Vijver, F. J. R., & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology, 33,* 141–156.

van Hemert, D. A., van de Vijver, F. J. R., Poortinga, Y. H., & Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences, 33,* 1229–1249.