The SAGE Handbook of Innovation in Social Research Methods

Edited by Malcolm Williams and W. Paul Vogt
Convergence of Structural Equation Modeling and Multilevel Modeling

Rex B. Kline

Everything is related to everything, but near things are more related than distant things.

The quote that opens this chapter is Waldo Tobler's First Law of Geography (quoted in Longley et al., 2005: 65), and it emphasizes that all places are more or less similar, but nearby places are more akin than distant places. This law could also describe the attainment of wisdom about some domain of study, that is, a sense of the big picture about how things that on the surface appear to be unrelated are actually connected on a more fundamental level. Deep learning is another term for this kind of understanding, and it is contrasted with surface learning, which involves more rote memorization than a conceptual reorganization that transfers to other situations and affects future problem solving (Lombardo, n.d.). Many have argued that the ultimate goal of education should be the facilitation of deep learning instead of just the accumulation of facts or situation-specific skills (e.g. Gardner, 1999).

It is also true that part of maturing as a researcher should involve an increasing sense of how many statistical techniques are, despite the use of different names or computer tools for the same basic analytical method, fundamentally related. An example was the realization starting in the 1960s that all forms of the analysis
of variance (ANOVA) are nothing more than a restricted case of multiple regression (e.g. J. Cohen, 1968), which itself is just an extension of bivariate regression that analyzes one or more predictors (factors, independent variables) of a continuous criterion (outcome, dependent variable). Any predictor in MR can be continuous or categorical, and both types of predictors can be analyzed together in the same equation. In contrast, it is awkward in ANOVA to include a continuous variable, such as age in years, as a factor in the analysis. One way to do so is to categorize a continuous variable into a dichotomy, such as through a median split, or into three or more discrete levels that make up an ordinal scale (e.g. low, medium, high). However, categorization – Thompson (2006) used the perhaps more apt term mutilation – of a continuous predictor is generally a bad idea for a few reasons. These include the loss of numerical information about individual differences in the original distribution and the possible introduction of statistically-significant-but-artifactual results due to mutilation (e.g. MacCallum et al., 2002). The analysis could be enhanced by the insight that categorization of continuous predictors is unnecessary when using a regression procedure instead of an ANOVA procedure.

Relations among behavioral science statistical methods extend even further. For instance, MR is part of the general linear model (GLM) family of techniques that include canonical correlation and multivariate analysis of variance (MANOVA) when there are multiple outcome variables and also some methods of exploratory factor analysis, such as principle components, when there is no distinction between predictor and outcome variables. All parametric GLM techniques are in turn subsumed under the family of techniques referred to as structural equation modeling (SEM), also known as covariance structure analysis (e.g. Fan, 1997). As a whole, SEM techniques are highly versatile and permit the evaluation of a wide range of hypotheses, such as those about direct or indirect effects, associations between manifest (observed) variables and latent variables (e.g. measurement), or means of observed or latent variables. In part due to their flexibility, SEM techniques have become quite popular among researchers. Indeed, it is increasingly hard to look through an issue of a research journal in the behavioral sciences and not find at least one article in which results of SEM analyses are reported.

A different family of statistical methods known as multilevel modeling (MLM) – also referred to as hierarchical linear modeling, random coefficients modeling, and mixed effects modeling, among other variations – has been emphasized in areas where hierarchical or clustered datasets are routinely analyzed. In such datasets, individuals (cases) are grouped into higher units, such as siblings within families or workers within departments. These larger units may themselves be nested under even higher-order variables, such as families within neighborhoods or departments within companies. Within each level, scores may not be independent. For example, siblings may all be affected by common family characteristics, such as total family income, and family characteristics may be affected by common neighborhood variables, such as community socioeconomic
status (SES). Repeated measures datasets are also hierarchical in that multiple scores are clustered under each case, and these scores may not be independent.

Many standard statistical techniques for single-level analyses where cases are not clustered, such as MR and ANOVA for between-subject effects, assume independence of the scores. Statistical tests in MR also generally assume that the residuals (error scores) are not only independent but also normally distributed and homoscedastic, or that their variance is constant across all levels of the criterion. Violation of these assumptions generally results in negatively biased estimates of standard errors, which means that they are on average too small. Because standard errors are the denominators of some statistical tests, such as the t-test, then the results of such tests could be statistically significant too often. That is, the actual Type I error rate (e.g. 0.55) could be much higher than the stated level of statistical significance (e.g. α = 0.05) when a statistical test that assumes independence is conducted with dependent scores. Estimates of certain population parameters, such as regression coefficients, could be biased too. With the exception of the analysis of repeated measures data, SEM also generally assumes independence. Overall, SEM is better suited for single-level analyses in datasets that are not hierarchical. However, it can be difficult to estimate both direct and indirect effects also effects of latent variables measured by multiple indicators in MLM, but these kinds of analyses pose no special problem in SEM. Thus, the SEM and MLM families of techniques have some complementary strengths and weaknesses.

Just as GLM and SEM can be viewed as extensions of basic principles of MR to the analysis of, respectively, multiple outcomes or latent variables, so too can MLM, but here core MR principles are extended to multilevel analysis of hierarchical data (e.g. Bickel, 2007). This common lineage also implies that SEM and MLM are more closely related to each other than might seem at first glance. For example, there are some specialized computer programs for MLM - including MLwiN 2.02 (Rasbash et al., 2005) and Hierarchical Linear and Nonlinear Modeling (HLM) 6.06 (Raudenbush et al., 2008), among others but they cannot generally analyze structural equation models with latent variables measured by multiple indicators.

Likewise, it has been awkward until recently to analyze multilevel models with specialized software for SEM, including older versions of LISREL, EQS, and Amos, among others. A notable exception is Mplus, which has been capable of SEM analyses and MLM analyses through its last several versions, the most recent of which is Mplus 5 (L. Muthén and B. Muthén, 1998–2007). Indeed, the analysis of multilevel structural equation models with latent variables measured by multiple indicators is quite straightforward in Mplus. Also, the most recent versions of LISREL (8.8; Jöreskog and Sörbom, 2006) and EQS (6.1; Bentler, 2005) include special syntax and features for analyzing multilevel structural equation models.

Work published over the last two decades by several different authors about commonalities between SEM and MLM has facilitated the development of the computer tools just mentioned. For example, B. Muthén (1994) described how to
represent and estimate multilevel confirmatory factor analysis models using (then) standard SEM software and notation. This approach involved exploiting the capability of SEM computer tools to analyze models across multiple groups, but in this case the 'groups' corresponded to within-group variation versus between-group variation in the same hierarchical dataset. McArdle and Hamagami (1996) described how different kinds of multilevel models could be specified as instances of multiple-group structural equation models. More recently, Bauer (2003) and Curran (2003) demonstrated how structural equation models and multilevel models are analytically equivalent for certain kinds of hierarchical datasets. See the two works just mentioned for additional citations of important work about the convergence of SEM and MLM.

A more complete synthesis of the two sets of techniques is described as multilevel structural equation modeling (ML-SEM) (e.g. Heck, 2001; Kaplan, 2000: Chapter 7; Rabe-Hesketh et al., 2007), and it offers potential advantages to researchers who are familiar with either SEM or MLM, but not both. Considered next are the basic rationales of MLM and SEM with emphasis on their complementary strengths and weaknesses. After review of these foundational issues, examples of the kinds of hypotheses and models that can be tested in ML-SEM are considered.

RATIONALE OF MULTILEVEL MODELING

There are two main contexts for the analysis of hierarchical data. The first is in repeated measures designs where scores are clustered under each case. Expected dependence among such scores is accounted for in some standard statistical techniques. For example, the error term in repeated measures ANOVA takes account of score covariances across the levels of repeated measures factors. However, ANOVA assumes that the error variances of repeated measures variables are equal and independent. These restrictive assumptions are often violated in actual repeated measures datasets. The technique of MANOVA can also be applied to repeated measures data, and its assumptions about error variance are less restrictive (e.g. errors may covary). A special strength of SEM is that it allows even greater flexibility in the modeling of error covariances compared with ANOVA or MANOVA, a point that is elaborated later.

The second context is the use of a complex sampling design in which the levels of at least one higher-order variable are selected prior to sampling of individual cases within each level. An example is the method of cluster sampling. Suppose in a study of Grade 2 scholastic skills a total of 100 public elementary schools in a particular geographic region is randomly selected, and then every Grade 2 student in these schools is assessed. Here, students are clustered within schools. A variation for this example is multistage sampling where only a portion of the students within each school are randomly selected (e.g. 10%) for inclusion in the sample. In stratified sampling, a population is divided into homogenous, mutually exclusive subpopulations (strata), such as by gender or ethnic categories, and
then cases within each stratum are randomly selected. The resulting hierarchical dataset may be representative of the variable(s) selected for stratification.

In a complex sample, scores within each level of a higher-order variable may have some degree of dependence. This means that the application of standard formulas for estimating standard errors that assume independence may not yield correct results. These formulas may be reasonably accurate in a single-level sample, but they tend to underestimate sampling variance for dependent scores. Thus, one motivation for the development of multilevel statistical methods is the need to correctly estimate standard errors in complex sampling designs.

A second motivation for multilevel analysis is the study of effects of contextual variables on scores at lower levels in a hierarchical dataset. Suppose that a researcher wishes to study factors that predict scholastic achievement among Grade 2 students. The researcher will measure characteristics of students, such as their gender, ethnicity, and family income. In a complex sampling design, the students will be selected from within a total of 100 different schools. Characteristics of the students’ schools will also be measured, such as their size (total student body) and degree of emphasis on academic excellence. The variables just mentioned are uniquely school-level characteristics.

However, it is also possible to aggregate selected student-level variables up to the school level and to consider these aggregated variables as school-level variables, too. For example, gender is a dichotomous variable for individual students, but the total proportion of students who are girls at each school is a school-level characteristic. Likewise, the average family income over all students who attend the same school is a characteristic of that school. Measured across all schools, the proportion of girl students or average family income could be between-school predictors of student achievement at the individual level. In a multilevel statistical analysis, both student-level and school-level predictors of achievement could be analyzed together in a way that (1) correctly estimates standard errors and (2) simultaneously incorporates data from the two different levels, within groups and between groups. These features lead to (3) separate estimates of between-group effects (e.g. school size) and within-groups effects (e.g. student gender) on variables of interest (e.g. achievement).

Often the first step in analyzing a hierarchical dataset is the calculation of a statistical index of the degree to which observations at the case level, such as students, depend on a grouping or cluster variable, such as schools. One such index is the unconditional intraclass correlation coefficient (UICC), and its value indicates the proportion of total variability explained by the grouping variable. One way to estimate the UICC is to conduct a standard, fixed-effects ANOVA where the grouping variable is the single factor and scores at the case level are the dependent variable. From the results of this analysis, UICC is calculated as

\[
\hat{\rho} = \frac{MS_c - MS_w}{MS_c + (df_c)MS_v}
\]

(1)
where \( \hat{\rho} \) is the sample estimate of the UICC, \( MS_c \) is the between-groups mean square for the cluster variable (e.g. schools), \( MS_w \) is the pooled-within groups mean square (e.g. students within schools), and \( df_c \) is the degree of freedom for the grouping variable, which is one less than the number of clusters. An alternative is to conduct a random-effects (variance components) ANOVA where the grouping variable is specified as a random factor:

\[
\hat{\rho} = \frac{\hat{\sigma}_c^2}{\hat{\sigma}_c^2 + \hat{\sigma}_e^2}
\]  

(2)

where \( \hat{\sigma}_c^2 \) is the variance component estimate for the grouping variable and \( \hat{\sigma}_e^2 \) is the variance component estimate for error. In SPSS, the value of \( \hat{\rho} \) estimated using either equation (1) or (2) within the same hierarchical dataset should be equivalent within rounding error when the estimation method is specified in the Variance Components procedure as ‘ANOVA’; otherwise, the two values may be somewhat different.

If the value of \( \hat{\rho} \) is high enough, then it may be necessary to employ MLM instead of a standard, single-level statistical technique. Unfortunately, there is no ‘gold standard’ concerning cut-off values for \( \hat{\rho} \) above which would clearly indicate a problem if MLM were not used. Some sources suggest that UICC values as low as 0.10 may be sufficient to result in appreciable bias in standard errors if MLM techniques are not used (e.g. Maas and Hox, 2005). Another standard is that MLM should be conducted if the value of \( \hat{\rho} \) is statistically significant, but it is possible that near-zero values of \( \hat{\rho} \), such as \( \hat{\rho} = 0.01 \), could be statistically significant in a large sample. In any event, the decision rule ‘if \( \hat{\rho} \geq 0.10 \), then use multilevel modeling’ may be a relatively conservative one; see Bickel (2007: ch. 3) for more information.

An example of the need to separately estimate effects at different levels of analysis described by Stapleton (2006) is considered next. Suppose that a researcher believes that students who spend more time watching television (TV) have lower levels of scholastic achievement (Ach). Both variables are measured among students who attend four different schools; that is, students are clustered within schools. The scatterplots for each of four hypothetical schools are presented in Figure 26.1. The group centroid – the point that represents the mean on both variables – and the within-group regression line are represented in the figure for each scatterplot with, respectively, a dot or a dashed line. Within each school, the association between TV and Ach is negative; that is, more time spent watching television predicts lower achievement. From the perspective of SEM, the same basic within-group covariance structure holds across schools.

However, there is another aspect of the relation between the variables TV and Ach in Figure 26.1 that is apparent from a between-school perspective: There is a positive association between the average number of hours of television watched at a school and the average achievement in a school. This positive covariance is apparent if you draw a line in the figure that connects the four group centroids. The fact that the within-school versus between-school associations between the
variables TV and Ach are of opposite signs (respectively, negative, positive) is not contradictory. This is because the between-school association is estimated using group statistics (means), but the within-school associations are estimated using scores from individual students.

In Figure 26.1, the within-school slopes are identical, but the within-school intercepts differ across the schools. This particular pattern may be improbable. Specifically, given a more realistic number of schools in a two-level dataset, such as about 100 or so, it may be more likely that (1) both the slopes and the intercepts of the within-school regression lines may vary across schools. Furthermore, it is also possible that (2) part of the variability in slopes or intercepts is explained by at least one contextual variable, such as school size. For example, the relation between television watching and achievement could be stronger in smaller schools but weaker in larger schools, or even close to zero. It is also possible that (3) variability in slopes is related to variability in intercepts, both across schools. For example, a higher school average level of achievement (i.e., a higher intercept) may predict a stronger negative relation between television watching and achievement, and vice-versa. That is, the covariance between intercepts and slopes across schools may not be zero.

**Random coefficient regression**

Perhaps the most basic multilevel statistical technique is random coefficient regression. Unlike in standard MR for single-level datasets where slopes and intercepts are conceptualized as fixed population parameters, coefficients for slopes and intercepts in random coefficient regression can be specified as random
effects that vary (and covary) across subpopulations. In a two-level dataset, these subpopulations correspond to levels of cluster variables, such as schools for the example discussed to this point about the relation between television watching and achievement. Specifically, the researcher could specify in a random coefficient regression that both the within-school slopes and intercepts vary across schools. In contrast to a standard MR analysis where separate sets of slopes and intercepts would be estimated for each and every school, two different types of parameters are estimated in random coefficient regression. These include (1) the variance and covariance of the slopes and intercepts, and (2) the weighted average slope and intercept across all schools. The latter are the fixed components of the randomly-varying slopes and intercepts, and the former are the random components of these parameters.

A related point is that ordinary least squares (OLS) is not the typical estimation method used in random coefficient regression as it is in standard MR. Briefly, if the cluster sizes at the second level are all equal (e.g. the size of every school is the same), then the design is balanced. In this case, it may be possible to use full-information maximum likelihood (FIML) estimation in a random coefficient regression analysis. This assumes that the number of clusters is large, say, > 75 or so (e.g. Maas and Hox, 2005) and also that the total number of cases across all clusters is large, too. In unbalanced designs, it may be necessary to use an approximate maximum likelihood estimator, one that is less computationally intensive but accommodates unequal cluster sizes. An example is restricted maximum likelihood (REML), which is available in the Linear Mixed Models procedure of SPSS for multilevel analyses.

**Multilevel regression**

In a basic random coefficient regression, there are no predictors of the variances and covariances among within-group slopes and intercepts across clusters. There are such predictors in a full multilevel regression analysis, and these predictors are typically contextual variables, or characteristics of groups. Presented next are the equations that specify a hypothetical two-level regression analysis where television watching and achievement are each measured within 100 different schools and where school size is a predictor of within-school regression coefficients. The level-1 within-school regression equation is

$$\text{Ach}_{ig} = \beta_{0g} + \beta_{1g} \text{TV}_{ig} + r_{ig}$$  \hspace{1cm} (3)

where $\text{Ach}_{ig}$ is the $i$th student's achievement score in the $g$th school and $\text{TV}_{ig}$ is the student level predictor of amount of television watching. The terms $\beta_{0g}$ and $\beta_{1g}$ represent, respectively, the random intercept and slope of the line for the regression of achievement on television watching for the $g$th school (e.g. Kaplan, 2000: 132–4).
Continuing with the same example, there are two equations at level-2, the between-school level, one for each of the random intercepts and slopes from equation (3):

\[
\beta_{0g} = \gamma_{00} + \gamma_{01} \text{Size}_g + u_{0g}
\]
\[
\beta_{1g} = \gamma_{10} + \gamma_{11} \text{Size}_g + u_{1g}
\]

where \(\text{Size}_g\) is the size of the \(g\)th school, \(\gamma_{00}\) is the mean intercept across all schools, and \(\gamma_{01}\) is the slope of the regression line for the relation between school mean achievement and the mean amount of television watching (e.g., see Figure 26.1). The terms \(\gamma_{00}\) and \(\gamma_{10}\) are the fixed components (averages) of the random intercepts and slopes. The terms \(u_{0g}\) and \(u_{1g}\) in equations (4) and (5) refer to, respectively, variability in the random intercepts and slopes; thus, they are the random components of the slopes and intercepts. The specification of intercepts and slopes as criterion variables in multilevel regression is described as an intercepts- and slopes-as-outcomes model.

Substituting the terms in the right-hand sides of equations (4) and (5) for, respectively, \(\beta_{0g}\) and \(\beta_{1g}\) back into equation (3) generates the full regression model at the student level, which is

\[
\text{Ach}_{tg} = \gamma_{00} + \gamma_{01} \text{Size}_g + \gamma_{10} \text{TV}_{tg} + \gamma_{11} \text{Size}_g \text{TV}_{tg} + (u_{0g} + u_{1g} \text{TV}_{tg} + r_{tg}).
\]

In words, the full model specifies that achievement of individual students is a function of the overall intercept (\(\gamma_{00}\)), the main effect of school size (\(\gamma_{01} \text{Size}_g\)), the main effect of the students’ time spent watching television (\(\gamma_{10} \text{TV}_{tg}\)), and the cross-level interaction effect where school size moderates the relation between watching television and achievement (\(\gamma_{11} \text{Size}_g \text{TV}_{tg}\)). The remaining terms in parentheses in equation (6) correspond to error variance from both levels of the model (see equations (3)-(5)).

Output from a computer tool for a multilevel regression analysis would include the regression coefficients for the terms specified in equation (6) and estimates of the residual (error) variance and also of the variances and covariances of the random slopes and intercepts. A two-level regression model can quickly become quite complicated as additional individual- or group-level predictors are added to the model. Also, the researcher must make many decisions about specification, including the designation of individual-level intercepts or slopes as random versus fixed effects or whether random components covary or not. The specification of covariance patterns among predictor or outcome variables is familiar to those who work mainly with SEM. Also compared with SEM, it is just as crucial in MLM to balance model complexity against parsimony and theoretical justification. That is, the goal in both techniques is to find the simplest model with a sound rationale that also fits the data.

A multilevel regression analysis can be expanded in many ways. For example, it is possible to estimate models with three or more levels (i.e., there are two
levels of cluster variables), but the complexity of such models can be daunting. This is in part explains why most multilevel models described in the research literature are two-level models. The analysis of repeated measures from a multilevel perspective also affords much flexibility. An example concerns latent growth models, in which the slopes and intercepts of repeated measures variables are treated as latent outcome variables that are predicted by either time-invariant variables (i.e., they are measured once) or other repeated measures variables. A standard textbook for MLM is Raudenbush and Bryk (2002), but its presentation is mathematically rigorous; see Bickel (2007) for a more-introductory level presentation that emphasizes the connection between MLM and standard OLS regression.

Summarized next are some limitations of MLM (Bauer, 2003; Curran, 2003):

1. Scores on individual- or group-level predictors in MLM are observed variables that are assumed to be perfectly reliable. (The same assumption applies to predictors in single-level MR, too.) This requirement is unrealistic, especially for predictors that are psychological variables measured with questionnaires instead of demographic variables.

2. There is no straightforward way in MLM to represent either predictor or outcome variables as latent variables (constructs) measured by multiple indicators. In other words, it is difficult to specify a measurement model as part of a multilevel model.

3. Although there are methods to estimate indirect effects apart from direct effects in MLM (e.g., Krull and MacKinnon, 2001), they can be difficult to apply in practice.

4. There are statistical tests of individual coefficients or of variances/covariances in multilevel regression, but there is no single inferential test of the model as a whole. Instead, the relative predictive power of alternative multilevel models estimated in the same sample can be evaluated in MLM (e.g., Bickel, 2007: ch. 3).

RATIONALE OF STRUCTURAL EQUATION MODELING

Diagrams of structural equation models are presented next using the reticular action modeling (RAM) symbol set (e.g., McArdle and McDonald, 1984). The RAM symbolism explicitly represents every model parameter that requires a statistical estimate.

Path models

A basic covariance structure consists of a structural model or a measurement model. Presented in Figure 26.2(a) is an example of a structural model for observed variables, or a path model. The observed variables in this model, $X_1$, $Y_1$, and $F_3$, are each represented with squares (rectangles can also represent observed variables). Each line with a single arrowhead ($\rightarrow$) in Figure 26.2(a) represents a hypothesized direct effect of one variable on another. (Direct effects are also called paths.) The arrowhead points to the presumed effect and the line originates from a presumed cause, such as $X_1 \rightarrow Y_1$ in the figure. Variable $X_1$ in this model is exogenous because whatever is presumed to cause this variable is not represented
in the path model. In contrast, variables \( Y_1 \) and \( Y_3 \) are endogenous because they are specified as outcomes of other observed variables.

Note in the path model of Figure 26.2(a) the specification \( X_1 \rightarrow Y_1 \rightarrow Y_3 \), which represents the presumed indirect effect of \( X_1 \) on \( Y_3 \) through a mediator variable, \( Y_1 \). This specification reflects the hypothesis that \( X_1 \) has an effect on \( Y_1 \) and that part of this effect is then 'transmitted' on to \( Y_3 \). Note that the path model in the figure also contains a direct effect of \( X_1 \) on \( Y_3 \), or \( X_1 \rightarrow Y_3 \). The separate estimation of direct versus indirect effects is a standard part of path analysis that is referred to as effects decomposition. The circles in the path model of Figure 26.2(a) represent latent variables (ellipses can also represent latent variables), in this case the disturbances of the endogenous variables. Disturbances represent all omitted causes of an endogenous variable, and they are considered in SEM as unmeasured exogenous variables. The numbers (1) that appear in the figure are scaling constants. These specifications assign a scale to each disturbance that is related to that of the unexplained variance of the corresponding endogenous variable. The two-headed curved arrows that exit and re-enter the same variable in the figure represent the variances of the exogenous variables, which are also considered model parameters because these variables are free to vary (and covary, too).

**Structural-regression models**

A limitation of path analysis is that scores on exogenous variables are assumed to be perfectly reliable. Also, measurement error in endogenous variables is reflected in their disturbances along with systematic variance not otherwise explained by predictors. That is, estimation of the proportion of unexplained variance is confounded with that of measurement error in path analysis. An alternative to these restrictive assumptions is to specify a structural-regression (SR) model, which has both a structural model and a measurement model. In the latter, observed variables are specified as multiple indicators of latent variables (constructs, factor). Direct and indirect effects are specified in the structural model, but in SR models these effects are between latent variables. This synthesis of regression analysis and factor analysis allows the separate estimation of measurement error in observed variables from unexplained variance in underlying factors.

The SR model presented in Figure 26.2(b) contains the same basic pattern of structural relations specified in the path model of Figure 26.2(a), but these effects involve latent variables in the SR model. Each latent variable in Figure 26.2(b) is specified as measured by either two or three observed variables, and the direct effects from factors to indicators (e.g. \( A \rightarrow X_2 \)) represent the influences of factors on scores. The scaling constants (1) that appear in the SR model next to the path for one indicator of each factor (e.g. \( A \rightarrow X_2 \)) assign a scale to each factor related to that of the explained variance of the corresponding indicator. Note in Figure 26.2(b) that (1) every indicator has a residual term that represents, in part, measurement error (unreliability); and (2) every endogenous factor (\( B, C \))
Figure 26.2 Examples of a path model (a) and a structural-regression model (b)
has a disturbance. Through this specification, it is possible to estimate direct and indirect effects between factors (e.g. $A \rightarrow B$), controlling for measurement error in their indicators.

**Confirmatory factor analysis models**

When analyzing an SR model, it is often important to first verify its measurement model, that is, to test the hypotheses about the correspondence between the indicators and the factors they are supposed to measure. If these hypotheses are wrong, then knowing relations among the factors specified in the structural part of the model may be of little value (Thompson, 2000). For example, if the three-factor measurement model implied by the SR model in Figure 26.2(b) did not explain covariance patterns among the seven indicators, then the fit of the whole SR model may be poor and the path coefficients may have little interpretive value. This is why many researchers use a two-step method to analyze SR models described by Anderson and Gebring (1988), as follows: In the first step, evaluate the measurement model implied by the original SR model. If this model is rejected, then it must be respecified. Given a satisfactory measurement model, the second step involves the testing of hypotheses about direct and indirect effects; that is, the structural part of the SR model is now analyzed.

Confirmatory factor analysis (CFA) is the SEM technique for evaluating pure measurement models. In such models, all associations between the factors are specified as unanalyzed, which implies that the factors covary, but we have no specific explanation about why they do so. An example of a two-factor CFA model is presented in Figure 26.3(a). In this model, the six observed variables are specified as measuring two factors, where indicators $X_1$-$X_3$ are presumed to reflect factor $A$ and indicators $X_4$-$X_6$ are presumed to reflect factor $B$. The curved line with two arrowheads in Figure 26.3(a) that connects the factors represents their covariance. If this path were replaced with a direct effect, such as $A \rightarrow B$, then the CFA model in Figure 26.3(a) would be transformed into an SR model.

**Latent growth models**

A latent growth model as specified in SEM is presented in Figure 26.3(b). In addition to its covariance structure, this model has a mean structure, which is represented in RAM symbolism by the graphical symbol $\triangle$, and here it represents a constant that equals 1 for every case. In SEM, this constant is treated as an exogenous variable (even though its variance is zero) that has direct or indirect effects on other variables in the model except for residual terms. The unstandardized coefficients for effects of the constant are interpreted as either means or intercepts. (This is the same basic method carried out 'behind the scenes' when a modern computer program for regression calculates an intercept.) In the model in Figure 26.3(b), these specifications result in the estimation of the means of the
two factors and also of the intercepts for the repeated measures outcome variable, which is represented as $Y_1 - Y_4$ in the figure.

These variables $Y_1 - Y_4$ in Figure 26.3(b) are specified as the indicators of two latent growth factors, Initial Status (IS) and Linear Change (LC). The IS factor represents the baseline level of variable $Y$, adjusted for measurement error. Because the IS factor is basically an intercept, loadings of $Y_1 - Y_4$ on IS are all specified as 1. In contrast, loadings on the LC factor are fixed to constants that correspond to the times of measurement, beginning with 0 for the first measurement and ending with 3 for the last. These constants (0, 1, 2, 3) specify a positive linear trend, but one that is adjusted for measurement error when the model is estimated.
In an actual analysis, one would obtain estimates of the factor means and also of their variances and covariances. For example, the mean for the LS factor in Figure 26.3(b) is the weighted average linear change (slope) over time across all cases. This mean is analogous to the fixed component of a random effect in MLM. The estimate of the variance of the LS factor indicates the range of individual difference in slopes across the cases. This variance is akin to the random component of a random effect in MLM. Also, the IS and LC factors in Figure 26.3(b) are specified to covary, and the estimate of this covariance indicates the degree to which initial levels of externalization predict rates of subsequent linear change, again corrected for measurement error.

The model in Figure 26.3(b) also has an error covariance structure, one where the error terms of the repeated measures indicators are allowed to covary across adjacent times (e.g. between \( E_{1} \) and \( E_{2} \)). Other patterns are possible, including no error covariances (i.e. the errors are independent over time) or the specification of additional error covariances (e.g. between \( E_{H} \) and \( E_{Y} \)). Finally, the model in Figure 26.3(b) includes a predictor, variable \( X \), of the IS and LC factors. This predictor could be either a continuous variable, such as family income, or a dichotomous one, such as gender. In this model, \( X \) is specified as a predictor of latent intercepts and slopes. The capability to specify observed variables as predictors of intercepts and slopes treated as latent variables in multilevel regression was mentioned earlier.

The analysis of latent growth models in SEM requires (1) a continuous outcome variable measured on at least three occasions; (2) scores that have the same units over time; and (3) data that are time structured, which means that cases are all tested at the same intervals. In contrast, the analysis of latent growth models in MLM does not require time-structured data, so it is even more flexible than SEM for analyzing latent growth models. As noted by Bauer (2003), Curran (2003), and others, latent growth models analyzed in SEM are in fact multilevel (two-level) models, ones that explicitly acknowledge the fact that scores are clustered under individuals (i.e. repeated measures). A latent growth model is specified differently in MLM – specifically, time is treated as a predictor in MLM, but time is represented in SEM via factor loadings that designate measurement occasions (e.g. the set 0, 1, 2, 3 for the LS factor in Figure 26.3(b)) – but SEM and MLM computer programs generate the same basic parameter estimates for the same latent growth model. This point of isomorphism between MLM and SEM is a major basis for relating the two techniques (Curran, 2003).

**Multiple-group analysis**

Essentially any type of structural equation model can be analyzed across multiple samples. The main question addressed in a multiple-sample SEM is whether values of estimated model parameters vary appreciably across the groups. If so, then (1) group membership moderates the relations specified in the model (i.e. there is a group \( \times \) model interaction), and (2) separate estimates of some model
parameters may be needed for each group. The capability to simultaneously estimate a model across multiple samples adds much flexibility to SEM. This is also another point of contact between SEM and MLM. Specifically, some types of multilevel models can be represented as instances of multiple-group SEM (McArdle and Hamagami, 1996); this point is elaborated in the next section.

Many of the special strengths of SEM correspond to limitations of MLM. For example, it is possible to represent latent variables measured with multiple indicators as either predictor or outcome variables in SEM. As a consequence of this specification, measurement error is controlled in the analysis. Likewise, the estimation of direct or indirect effects is relatively straightforward in SEM. Finally, there are inferential tests of the fit of an entire structural equation model to sample data. This test is based on the familiar model (familiar in SEM) chi-square statistic with degrees of freedom that equal the differences between the number of observations (sample covariances and means) and parameters that require statistical estimates. More parsimonious models have greater degrees of freedom. The same model chi-square statistic can also be used to test the relative fit of two nested models, in this case where one model is a subset of the other. Except when analyzing a particular class of latent growth models, SEM does not directly take account of clustering in a multilevel dataset.

EXTENDING STRUCTURAL EQUATION MODELING TO MULTILEVEL ANALYSES

Early efforts to extend SEM were based on ‘tricking’ standard computer SEM programs into analyzing two-level models. The trick is to exploit the capability of the software to simultaneously estimate a structural equation model across two groups. However, in this case the ‘groups’ corresponded to two different models, a within-group or level-1 model and a between-group or level-2 model, both analyzed in the same complex sample. The data matrix for the level-1 model is the pooled within-group covariance matrix based on the variation of scores from individual cases around group means, and for the level-2 model it is the between-group covariance matrix based on the variation of the group means around the grand means. Because older versions of most SEM computer programs had no built-in capabilities for analyzing clustered data, it was usually necessary to calculate these two data matrices using an external program such as SPSS. The two data matrices were then submitted to the SEM computer program as external files or were included as part of the syntax (command) file.

Presented in Figure 26.4(a) is the model diagram for a two-level regression analysis conducted by tricking an SEM computer program into estimating a two-level model. This model corresponds to the hypothetical data represented in Figure 26.1 where the within-school covariance between television watching and achievement is negative, but the between-school covariance is positive. In Figure 26.4(a), the observed variables TV and Ach are each specified as the
single indicator of a within-school factor and a between-school factor (e.g., Ach\(_w\) \(\rightarrow\) Ach, Ach\(_B\) \(\rightarrow\) Ach). The scaling constants for the within-school factors both equal 1, but for the between-school factors these constants equal the square root of the cluster \(n_c\), or the number of cases in each school. Group size is constant in a balanced design; otherwise, it is calculated as

\[
\frac{N^2 - \sum_{g=1}^{G} n_i^2}{N(G-1)}
\]

where \(N\) is the total number of cases across all groups and \(G\) is the number of groups. At each level of the model in Figure 26.4(a), the Ach factor is regressed on the TV factor (i.e. TV\(_w\) \(\rightarrow\) Ach\(_w\), TV\(_B\) \(\rightarrow\) Ach\(_B\)). This specification tells the computer to derive separate estimates of the within-school and between-school regression coefficients.

Listed in Table 26.1 is EQS syntax for analyzing the two-level regression model in Figure 26.4(a) as a two-group structural equation model. Although EQS 6.1 has special syntax for multilevel analyses, it is not used in this example. Instead, the syntax in the table indicates how to trick EQS into analyzing a two-level model as a two-group structural equation model. I assumed for this analysis, a balanced design where data are collected from \(n_c = 50\) students in each of \(100\) different schools (i.e. \(N = 5,000\)). I also assumed equal and negative slopes of the within-school regression lines across the schools, which is consistent with the scatterplots represented in Figure 26.1.

Listed in the top part of Table 26.1 is syntax that specifies the within-school part of the model in Figure 26.4(a). This syntax also defines the pooled within-school covariance matrix, in which the observed covariance between the variables TV and Ach is \(-7.1\) (see Table 26.1). Syntax for the between-school part of the model in Figure 26.4 is listed in the lower part of Table 26.1. Part of the trick of manual model set-up is to specify the within-school regression as part of the between-school model and to also impose equality constraints across the within- and between-school models on the corresponding parameter estimates. Finally, the between-school syntax in Table 26.1 defines the between-school model covariance matrix, in which the observed covariance between the variables TV and Ach is \(10.6\).

I submitted the syntax in Table 26.1 to EQS 6.1, and the analysis ran without problem. The model in Figure 26.4 perfectly fits the data because one regression coefficient is estimated between two variables at each level, so the total degrees of freedom are zero. However, what is more interesting here is that EQS calculates different coefficients for the regression of Ach on TV for the within-school versus between-school parts of the model. Specifically, the unstandardized (and standardized) values for the within-school and between-school models are, respectively, \(-0.710 (-0.502)\) and \(1.180 (0.914)\). These results are consistent with the data representation in Figure 26.1. See Stapleton (2006) for additional examples.
Table 26.1  EQS syntax for manual set-up of the two-level regression model in Figure 26.4(a)

```
/title
within-school model
/specifications
cases=5000; variables=2; matrix=covariance; groups=2;
/labels
V1=TV; V2=Ach; F1=TV_W; F2=Ach_W;
/equations
V1=F1; V2=F2;
F2=*F1+D2;
/variances
F1=; D2=;
/matrix
10.0
-7.1 20.0
/end
/title
between-school model
/specifications
cases=100; variables=2; matrix=cov; method=ml;
/labels
V1=TV; V2=Ach; F1=TV_W; F2=Ach_W; F3=TV_B; F4=Ach_B;
/equations
F2=*F1+D2; F4=*F3+D4;
V1=F1+7.07F3; V2=F2+7.07F4;
/variances
F1=; D2=; F3=; D4=;
/matrix
25.0
10.6 45.0
/constraints
(1,F1,F1)=(2,F1,F1); (1,D2,D2)=(2,D2,D2);
(1,F2,F1)=(2,F2,F1);
/end
```

Although it is interesting that SEM computer programs can be tricked into analyzing a two-level model, there is a cost in terms of complexity. For example, the model diagram in Figure 26.4(a) and the syntax in Table 26.1 are both rather complicated for what is basically a trivial two-level regression analysis where both the slopes and intercepts do not vary or covary. If we rely on the same trick to analyze more realistic – and interesting – multilevel models, then the degree of complexity can quickly become daunting. For example, Bauer (2003) and Curran (2003) described how to specify and analyze the intercepts- and slopes-as-outcomes model defined by equations (3)–(6) using standard SEM notation and
(a) Two-level regression analysis

(b) Intercepts- and slopes-as-outcomes model

Figure 26.4 Diagrams for manual set-up in a computer program for structural equation modeling for a two-level regression analysis (a) and an intercepts- and slopes-as-outcomes model (b)

software without MLM capabilities (i.e. a variation of the trick). For simplicity's sake, let us assume a balanced design where only four students are selected from each of 100 different schools and also that there are no missing data (i.e. $n_c = 4, N = 400$). Scores on the variables TV and Ach are collected from each student as level-1 variables. Also, the size of each school is measured as a
Presented in Figure 26.4(b) is the structural equation model diagram for the multilevel regression analysis just described. At first glance, the model in Figure 26.4(b) resembles the latent growth model in Figure 26.3(b). Both models just mentioned have mean structures, latent variables that correspond to intercept and slope terms, and a predictor of the two terms. However, there are two crucial differences that make the model in Figure 26.4(b) much more difficult to analyze: First, the model in Figure 26.4(b) holds for just four students in a particular school. The achievement scores of the four students are designated in the figure as Ach₁, Ach₂, and so on. Because the ordering of the four scores in the data file is arbitrary, the four error terms associated with each score are constrained to have equal variances. The four achievement scores in Figure 26.4(b) are specified to load on an intercept factor and also on a slope factor. The intercept and slope concern the corresponding terms from the regression of the achievement scores on the television watching variable, TV, within this particular school. In Figure 26.4(b), the higher-order variable school size is specified as a predictor of the within-school intercept and slope. All loadings of the achievement indicators on the intercept factor equal 1, but the loadings on the slope factor are fixed to equal the scores on the television watching variable for each student. For example, the loading TV₁ for the path Slope → Ach₁ in Figure 26.4(b) is the score on the television watching variable for the first student.

Second, in an analysis of the model in Figure 26.4(b) across all schools, the computer must apply individual factor loading matrices where 'individual' actually means 'classroom'. That is, the factor loading matrix is unique to each classroom, and the elements of this matrix for the slope factor contain the observed scores on the TV variable for the students in each classroom (Curran, 2003). Not all SEM computer programs allow the specification of individual factor loading matrices, but one that does is Mx (Neale et al., 2002), a freely-available matrix algebra processor and numerical optimizer that can also estimate the full range of structural equation models. Analysis of the model in Figure 26.4(b) becomes even more complicated if the cluster size is a more realistic number (e.g. at least 100 students per school), and it becomes more complex still, if the design is unbalanced with unequal numbers of students in each school. Clearly, trying to trick an SEM computer program to estimate even a relatively simple multilevel regression quickly ‘becomes a remarkably complex, tedious, and error-prone task’ (Curran, 2003: 557), in other words, a data-management nightmare.

Fortunately, more and more computer programs for SEM feature special syntax that makes it easier to specify and estimate multilevel models. This is because use of this special syntax automates much of the analysis so that the computer, and not you, does most of work when analyzing multilevel models. For example, listed in Table 26.2 is Mplus syntax for the multilevel regression model defined by equations (3)-(6). The raw data are contained in the external
Table 26.2  Mplus syntax for the multilevel regression model defined by equations (3)–(6)

| title: multilevel regression model          |
| data: file = school.dat;                    |
| variable: names = Ach TV School Size;       |
| within = TV;                               |
| between = Size;                            |
| cluster = School;                          |
| centering = grandmean (TV);                |
| analysis: type = twolevel random;           |
| model:                                     |
|   %within%                                 |
|     s | Ach ON TV;                              |
| %between%                                  |
|   Ach s ON Size;                           |
|   Ach WITH s;                              |

file called 'school.dat', and the four variables are Ach, TV, School (attended), and Size (total enrollment of school attended). Next, the Mplus syntax in Table 26.2 indicates that TV is the within-group or level-1 predictor, the cluster variable is School, and the between-school or level-2 predictor is Size. Grand-mean centering of the TV variable is specified, and the analysis type is specified as two-level with random coefficients (see the table). Mplus syntax for the within-school part of the model defines the slope, labeled 's', from the regression of Ach on TV as a random variable. Syntax for the between-school part of the model listed in Table 26.2 specifies that the random slopes and random intercepts are regressed on the school size variable and that the random terms just mentioned covary. That's all there is to it. In general, Mplus syntax for both SEM and MLM is concise and straightforward.

Presented in Figure 26.5 are two examples of how diagrams for multilevel models are represented in the Mplus manual (L.K. Muthén and B.O. Muthén, 1998–2010). Compared with RAM symbolism, these abbreviated diagrams are simpler and relatively easy to understand. For example, the model in Figure 26.5(a) is an abbreviated version of the standard SEM model in Figure 26.4(a) for the basic two-level regression analysis where Ach is regressed on TV both within schools and between schools. The model diagram in Figure 26.5(a) indicates that the same basic regression analysis is conducted at both levels. Residual variance in this model is represented by the lines with single arrowheads oriented at 45° angles that point to outcome variables. The model in Figure 26.5(b) is an abbreviated version of the standard SEM model in Figure 26.4(b) for the intercepts- and slopes-as-outcome model defined by equations (3)–(6). In the within part of the model in Figure 26.5(b), the filled circle at the end of the arrow from TV to Ach represents a random intercept, and the filled circle on the arrow from TV to Ach labeled 's' represents a random slope. In the between part of the model in Figure 26.5(b), the random intercept is referred to as Ach, and it appears as a circle because it is conceptualized as a continuous latent variable that varies
CONVERGENCE OF SEM AND MLM

(a) Two-level regression analysis

(b) Intercepts- and slopes-as-outcomes model

Figure 26.5 Abbreviated diagrams for a two-level regression analysis (a) and an intercepts- and slopes-as-outcomes model (b)

across schools. Also represented in the between part of the model is (1) size as a predictor of random intercepts and random slopes, and (2) a presumed covariance between the two random terms, Ach (intercepts) and 's' (slopes). A mean structure is implied in Figure 26.5(b) because intercepts are estimated, but it is not explicitly depicted as in RAM notation.

MULTILEVEL STRUCTURAL EQUATION MODELING

Multilevel structural equation models are usually estimated in three basic steps. The first involves calculation of UICC values for outcome variables across the levels of the cluster variable. If more than about 10% of the variance is explained by between-group variation, then the need for multilevel structural equation modeling (ML-SEM) may be indicated. The next two steps parallel those of two-step estimation of SR models in SEM, but for ML-SEM these steps correspond to analysis of the within-group model only prior to simultaneous estimation of the within- and between-group models. The goal is to distinguish specification error
in either level, within versus between. Specifically, the within model is analyzed using just the pooled-within group covariances and means (i.e. the cluster variable is ignored). Although the fit of the within model may not be satisfactory due to the omission of between-group effects, the basic parameter estimates should nevertheless make sense. Next, the between model is specified, and then both the within and between models are simultaneously estimated using both the pooled within-group and between-group covariances and means. If the model is an intercepts- and slopes-as-outcomes-model, then between-group predictors of these random effects are added at the second step. Otherwise, one could specify the basic same model at the between level as for the within level. Poor model fit in this analysis may indicate the need to specify different models at within versus between levels. See Hox (2002) and Stapleton (2006) for description of additional possible analytic steps. Considered next are examples of ML-SEM.

This example of a multilevel path analysis (ML-PA) concerns a recent study by Wu (2007), who administered within a sample of 333 undergraduate students questionnaires about life satisfaction, what respondents say they want, or amount, and also of the gap between what they have and what they want, or have–want discrepancy. These questionnaires concerned 12 different life facets, including social support, health, physical safety, financial resources, and so on. Because ratings about life facets are repeated measures, Wu conceptualized that facets are nested under individuals. That is, the within-individual level concerns variation among facets for each respondent, and the total number of observations at this level equals 333 x 12, or 3,996. The between-individual level refers to differences across people that may affect ratings at the facet level, and the total number of observations at this level is 333. Values of the UICC across the 12 life facets ranged from 0.18–0.23, which indicate that about 18–23% of variance in life facet ratings is explained by between-individual variation.

At the both within and between levels, Wu hypothesized that both have–want discrepancy and amount have direct effects on life satisfaction and also that the former has an indirect effect through the latter (i.e. Discrepancy → Amount → Satisfaction). The structural model of the presumed direct and effects among these observed variables at the within level is presented in the top part of Figure 26.6(a). However, results of subsequent analyses indicated that a simpler structural model, one with no direct effect from amount to satisfaction, held at the between level, which is presented in the bottom part of Figure 26.6(a). That is, the effect of amount on satisfaction was mediated entirely by have–want discrepancy at the between-individual level, but not at the within-individual level where amount also directly affected satisfaction. Wu interpreted these results as suggesting that life satisfaction involves an explicit have–want comparison, but whether its effect is entirely indirect through its prior impact on overall amount of wants or not depends on the level of analysis, within- versus between-individuals. See Heck (2001) for description of an ML-PA where student characteristics, such as gender and minority status, and school characteristics, such as the degree of school quality, were specified as predictors of math skills
(a) Two-level path analysis

Discrepancy

Amount

Within

Between

Discrepancy

Amount

Satisfaction

(b) Two-level confirmatory factor analysis

\[ \text{It}\, 7 \quad \text{It}\, 8 \quad \text{It}\, 9 \quad \text{It}\, 10 \quad \text{It}\, 11 \quad \text{It}\, 12 \quad \text{It}\, 13 \quad \text{It}\, 14 \quad \text{It}\, 15 \]

Teacher Quality

Neg. Sch. Climate

Misbehav.

Within

Between

\[ \text{It}\, 7 \quad \text{It}\, 8 \quad \text{It}\, 9 \quad \text{It}\, 10 \quad \text{It}\, 11 \quad \text{It}\, 12 \quad \text{It}\, 13 \quad \text{It}\, 14 \quad \text{It}\, 15 \]

General Climate

Figure 26.6 A multilevel path analysis model (a) and a multilevel confirmatory factor analysis model (b)

among Grade 8 students in, respectively, a within-school path model versus a between-school path model.

The next example concerns the multilevel analysis of a measurement model, that is, one with latent variables measured by multiple indicators. Using data from a sample of over 10,000 public school Grade 10 students who attended about 1,000 different schools in the United States, Kaplan (2000: 48–53) described the results of a single-level CFA of a 15-item questionnaire about perceptions of school climate. Items included those about perceptions of teacher quality (e.g. teachers seen as interested in students), negative school environment
Kaplan found that a measurement model where 11 items were specified as indicators of three factors that corresponded to the three areas just mentioned had adequate fit to the data in the whole sample (i.e. ignoring school as a cluster variable).

In a subsequent analysis, Kaplan (pp. 136-40) conducted a multilevel confirmatory factor analysis (ML-CFA) in which the fact that students are nested under schools was explicitly represented. The ML-CFA model estimated by Kaplan is presented in Figure 26.6(b). The within-school CFA model in the figure is the same basic three-factor measurement model estimated by Kaplan in the single-level CFA. Names of the indicators in this model refer to questionnaire items (e.g. 'It?' in the figure means item no. 7). However, the measurement model at the between-school level in Figure 26.6(a) is a single-factor model where all indicators load one a general school climate factor. That is, this ML-CFA model reflects the hypothesis that within-school variation in student ratings is differentiated along three dimensions, but one general climate factor explains between school variation. This model had adequate fit to the hierarchical data for this analysis. See Dyer et al. (2005) for an example of an ML-CFA where organizational and national variables are included in models of the factor structure of leadership.

There are other many variations of ML-SEMs, including multilevel SR models, multilevel latent growth models, and so on. It is also possible to estimate intercepts- and slopes-as-outcome models where either predictor or outcome variables are latent variables measured with multiple indicators. An example of such a model is presented in Figure 26.7. The within model consists of a path model where the slopes and intercepts associated with all three paths are specified as random effects. In the between model, these slopes and intercepts are specified as outcome variables where the group-level predictor is a factor (A) measured by multiple indicators (X_2 - X_4), which controls for measurement error at this level. The whole multilevel model in Figure 26.7 reflects the hypotheses that (1) the magnitudes of the intercepts and slopes associated with the direct effects at the individual level vary as a function of the group variable, and (2) the intercepts and slopes from each direct effect are pairwise correlated. This whole model in Figure 26.7 features the representation of indirect effects (within model) and also takes account of measurement error (between model), a combination relatively unique to the convergence of SEM and MLM.

CONCLUSION

Researchers who know something about both SEM and MLM can test an even wider range of hypotheses compared with those who know about one technique, but not the other. Specifically, the convergence of the two techniques in the form of ML-SEM offers the ability to (1) calculate correct standard errors in
Within

Between

hierarchical datasets, (2) enter predictors from both the individual level and the group level (contextual effects) in the same analysis, (3) take account of unreliability when latent variables are represented as measured by multiple indicators, and (4) estimate both direct and indirect effects when structural models are analyzed, perhaps all in the same analysis depending on the model. The increasing availability of computer tools that directly support the analysis of multilevel structural equation models analyzed in complex samples is making it easier for researchers to actually reap these potential benefits. The costs for this increased flexibility in hypothesis testing include the need to make informed decisions about an even larger number of specification issues in ML-SEM compared with using either SEM alone or MLM alone. If these decisions are poor, then the results of ML-SEM may have little value, but this is true for any statistical modeling technique. Given good ideas based on relevant theory and results of empirical studies, though, it may be possible to better represent them in ML-SEM compared with either SEM or MLM alone.
NOTES

1 Code for conducting certain types of multilevel analyses in R, a free software environment for statistical computing and graphics production, is available in works such as de Leeuw (2008).
2 There are also examples of R code for SEM analyses; see Fox (2006).

REFERENCES


