

Applied Psychometrics: The Modeling Possibilities of Multilevel Confirmatory Factor Analysis (MLV CFA)

Theodoros A. Kyriazos

Psychology, Panteion University, Athens, Greece
Email: th.kyriazos@gmail.com

How to cite this paper: Kyriazos, T. A. (2019). Applied Psychometrics: The Modeling Possibilities of Multilevel Confirmatory Factor Analysis (MLV CFA). *Psychology*, 10, 777-798.
<https://doi.org/10.4236/psych.2019.106051>

Received: April 1, 2019

Accepted: May 6, 2019

Published: May 9, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Multilevel CFA models (MLV CFA) modeling permits more sophisticated construct validity research by examining relationships among factor structures, factor loadings, and errors at different hierarchical levels. In the MLV CFA models, the latent variable or variables have two kinds of elements: 1) the between-group elements (Level 2 or higher level) and 2) the within-group elements (Level 1 of lower level). The between-group elements represent the general part of the model and the within-group element the individual part. The within-level variation includes an individual-level measurement error variance, which generally expands the impact of the within-level variation to the intraclass correlations. Multilevel CFA therefore generates results corresponding to those generated by perfectly reliable measures. If the same measurement model is specified across levels, by defining each item loading to be invariant with its across-level counterpart, the researcher can equate the factor scales across levels. Thus, the factor variances at different levels are directly comparable. The fit of this constrained MLV CFA model can be evaluated by comparing it with an unconstrained model specified with freely estimated factor loadings at each level. In the present work the steps of the above procedure are fully described and additional issues relevant to the use of MLV CFA are discussed in detail.

Keywords

Multilevel Confirmatory Factor Analysis, MLV CFA, ML CFA, Hierarchical Factor Structure, Multilevel CFA Model, Between Groups, Within Groups, Intraclass Correlation Coefficient (ICC)

1. Introduction

In all the analyses carried out in the social and behavioral research field, data are

organized at a single level. Nevertheless, real world data are frequently structured in multiple levels. These data structures are called hierarchical¹. Such hierarchical structures are also termed nested data or clustered data (Byrne, 2012). This means that some variables are clustered or nested within other variables (Field, 2013; Geiser, 2013; Nezlek, 2011). For example, to study the attachment type of a child to its mother (Bowlby, 1969, 1973; Ainsworth, 1978), a researcher studies the mother-infant relationship of 300 infants in 100 families. The infants are nested in families. Infants are the first level of analysis and families are the second level. Lower-level units are also called *micro level* and higher-level *macro level* units (Heck & Thomas, 2015; Geiser, 2013). Macro-level variables are alternatively called groups or contexts (Kreft & de Leeuw, 1998; Heck & Thomas, 2015). Infants grew up in different family environments. Therefore, the researcher expects that they will have different attachment types. Generally, research in psychology deals with designs about individuals acting within a context, like families in the previous example, or schools (see Byrne, 2012; Geiser, 2013), organizations (see Brown, 2015; Darlington & Hayes, 2017) or neighborhoods (see Tabachnick & Fidell, 2013). The family in the above example is a contextual variable (Field, 2013) that multilevel modeling analysis allows to be taken into consideration (Hox, 2013; Loehlin & Beaujean, 2017). Models used to analyze clustered data are called Multilevel Models, Hierarchical Linear Models, Random Coefficient Models, or Mixed Models (Geiser, 2013; Field, 2013 among many others). Multilevel models are not a new conceptualization (cf. hierarchical linear models; Raudenbush & Bryk, 2002; Bickel, 2007; as quoted by Brown, 2015). However, only in recent decades, they were efficiently incorporated in CFA (c.f. Muthén, 1994, 1997, 2004; Brown, 2015).

The purpose of this study is to describe the procedure of Multilevel Confirmatory Factor Analysis Modeling (MLV CFA, Byrne, 2012), i.e., how to incorporate the multilevel approach into a CFA model.

2. Overview of Multilevel Modeling

Multi-level models are a category of statistical techniques for studying hierarchically structured data-sets where the scores 1) are nested into larger units (clusters) and 2) each cluster may be dependent from the other. For example, repeated measurements generate inherently hierarchical datasets with multiple scores clustered within each respondent (Kline, 2016: p. 444). Similarly, Selig, Card, and Little (2008) commented—as reproduced by Byrne (2012)—that any model representable as a multigroup SEM, can also be specified as a multilevel SEM/CFA, if the data are hierarchically clustered.

This relative delay of MLV CFA modeling could be attributed to the powerlessness of the older CFA software packages to deal with the inherent complexities of MLV CFA effectively, e.g. with the computation of separate covariance

¹This hierarchy is not identical to the higher order EFA and CFA. In FA the hierarchical element emerges from the construct been evaluated. Here the hierarchical element emerges from the data sampling design.

matrices for sampling units and the use of robust estimators (Heck & Thomas, 2009; Hox, 2002; McArdle & Hamagami, 1996 as quoted by Byrne, 2012). In MLV EFA and MLV CFA, both direct and indirect effects are considered simultaneously before the assessment of the overall model fit, thus they are very flexible (Hox, 2013). For a comparison of the multilevel design to the cross-sectional design see Table 1. For a conceptual representation of a Multilevel family functioning model see Figure 1.

A two-level structure (like in Figure 1) is the simplest hierarchy available (Field, 2013; Kline, 2016), i.e., at least one higher-level variable is included above individual cases (Kline, 2016). So, if in a study of family functioning the researcher decides to study the neighborhood within the families are nested, then a second-level hierarchical model must be constructed in which neighborhood is the second level. In a similar vein if multiple neighborhoods are included in the sample, then the area could become the third level (see Tabachnick & Fidell, 2013 for a similar example). Different neighborhoods and areas are contextual

Table 1. Conceptual differences of the multilevel research design and the crossed research design.

Multilevel Design				Crossed Design			
Parents (Level 2)	A		B	Parents (Level 2)	A		B
Children (Level 1)	1	2	3 4	Children (Level 1)	1	2	3 4 1 2 3 4

Source. Adapted by Schumacker & Lomax, 2016, page 195-196.

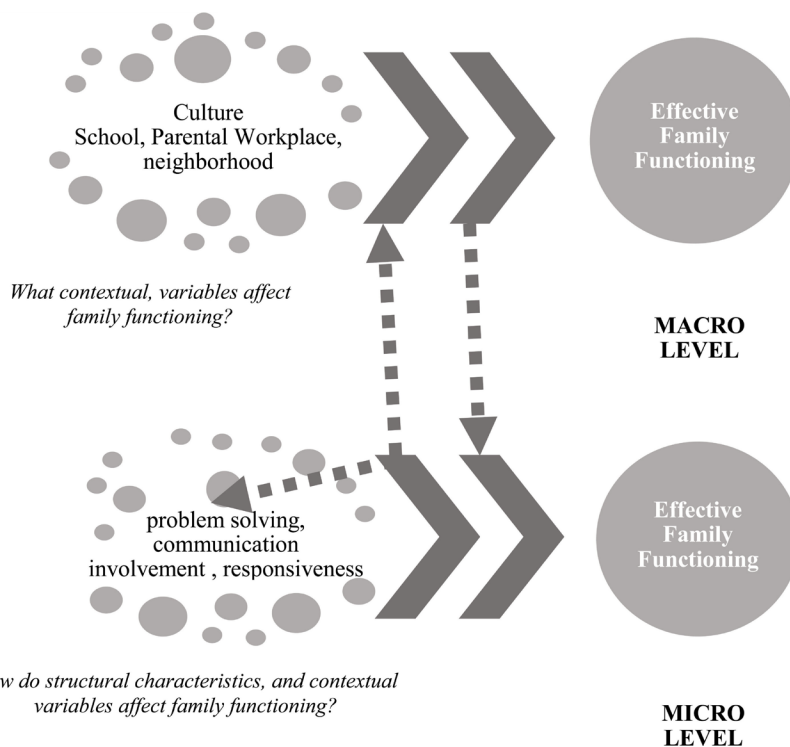


Figure 1. Conceptual multilevel representation of the McMaster model of family functioning (Epstein, Bishop & Levin, 1978).

variables, possibly reflecting differences in social and economic status or/and culture (see also Field, 2013 and Byrne, 2012 for analogous examples).

To use another common example (e.g., Field, 2013; Kline, 2016; Hox, 2013; Geiser, 2013), let us assume that a sample includes 3000 students who attend 20 different schools. Scores from students (1st level) attending the same classroom (2nd level) may also not be independent and scores from students enrolled in the same school (3rd level) may not be independent as well. This is likely to happen because students of the same classroom are affected by similar influences like the teacher's character and peer's behavior. According to Kline (2016) students of the same school could equally be influenced by school staff, school discipline frameworks, curriculums established, the number and nature of midterm exams and the like. Depending on the sampling design, there could be additional higher levels, e.g., schools, districts, cities, states, countries (Geiser, 2013).

This case similarity because of common contextual influences in clustered sampling is problematic because it violates two core assumptions of quantitative measurement: 1) that all cases are independent, and 2) that all random errors of cases are also independent, normally distributed, and homoscedastic (Byrne, 2012). By definition these assumptions are made by traditional statistical approaches like Ordinary Least Squares (OLS) regression analysis and Analysis of Variance (Cohen, Cohen, West, & Aikem, 2003; Geiser, 2013; Field, 2013). Therefore, using conventional statistical approaches to analyze clustered data may lead to biased results (Geiser, 2013). Specifically, a vital reason for MLV use is the correct estimation of standard errors or the assignment of probability weights in complex sampling designs (Kline, 2016; Kelloway, 2015; Brown, 2015; Field, 2013). The bias, Kline (2016) continues, arises because standard errors are denominators of significance tests, and when underestimated the null hypothesis is often rejected, as p values of the statistical significance tests will often be too small (Geiser, 2013). Moreover, clustering can lead to overestimation of the effective sample size. This would introduce biased statistical inference by an increase in the alpha error rate (Cohen et al., 2003; Snijders & Bosker, 1999; Geiser, 2013). Muthén and Satorra (1995) argue that the more similar the individuals within groups are, the more biased the parameter estimates, standard errors, and related tests for significance will emerge (as reproduced by Byrne, 2012).

An additional reason, multilevel structure should not be overlooked is that interactions of variables at different levels are often of central research interest (Geiser, 2013), see Figure 1 for a conceptual representation of interactions across levels. This is especially true for estimating the contextual effects of higher level variables on the scores of first-level cases, i.e., to examine within-and between-cluster relationships (Brown, 2015). For instance, returning into a previous example, schools differ in the number of enrolled students. This is a characteristic of schools, not of students (Kline, 2016; Geiser, 2013). That is, a macro level (level 2) variable affected the relationship of level-1 variables. This is called a *cross-level interaction* (Darlington & Hayes, 2017). The basic terms used

in MLV SEM and MLV CFA are presented in **Table 2**. See **Figure 2** for an example of hierarchical structure of families nested into cultures.

Although multilevel modeling was introduced to study individuals within groups, the method was extended to repeated measures data (like in longitudinal designs). Thus, measurement occasions (termed also time points) are nested within individuals (Bryk & Raudenbush, 1987; Goldstein 1987; Singer & Willett, 2003; Geiser, 2013). Multilevel modeling of longitudinal data is a powerful approach, because it offers many possibilities for the metric treatment of time points, dealing effectively with missing data from dropouts and panel attrition (Hox, 2013). Crucially, Structural Equation Models is more flexible approach than the traditional multilevel regression models additionally because regression models are based on unrealistic assumptions, e.g. that predictor variables are perfectly reliable. Structural equation models do not assume perfect reliability of variables, because they can specify a measurement model for the predictor or

Table 2. Terms used in multilevel SEM and SEM analysis.

Between groups	Model for the group-level structure (second level). This term gets progressively rather vague as higher levels are added, it is suggested to be combined with an explicit indication of the grouping element of each level (i.e., class or school level).
Cross-level interaction	Higher level variables can directly influence lower level variables. This is usually displayed by an interaction between higher-level and lower-level variables.
Fixed effect, fixed Coefficient	Factor loadings and path coefficients that do not change across the units of the higher level.
Intraclass Correlation	Used to examine population similarity of the individuals of the same group. It is also a measure of the amount of population variance at the group level.
Multilevel model	A model comprising variables at multiple levels of a hierarchically structured population. Also called hierarchical model.
Random effect, random coefficient	Factor loadings and path coefficients that do not change across the units of the higher level.
Variance Component	Variances and covariances of the changeable coefficients.
Within groups	Model representing the structure at the lowest level, i.e. the individual (first) level.

Source. Hox (2013: p. 292).

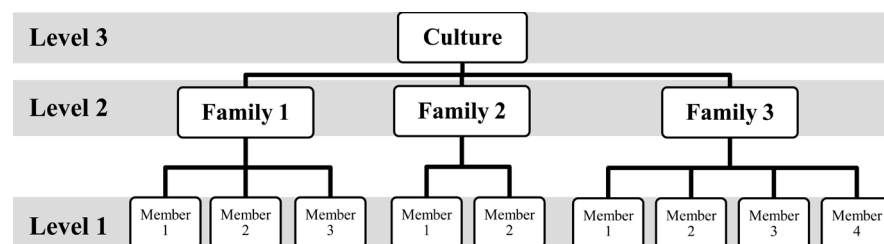


Figure 2. A 3 level hierarchy where respondents are nested within families and families within cultures (hierarchy is adapted from Byrne, 2012).

Table 3. Summary of the reasons why to use MLV modeling.

- Multilevel models properly account for the hierarchical data structure causing data dependencies.
- Multilevel modeling methodology overcomes standard error bias due to clustering that generating inflated Type-I error rates and inaccurate confidence intervals.
- Multilevel models permit analyzing variables at different levels taking into account cross-level interactions.
- Multilevel analysis is a more flexible method requiring fewer assumptions than other statistical methods such as, repeated measures of A NOVA.

Source. Geiser, 2013, page 197.

outcome variables. Additionally, they can model more complicated interactions, like indirect effects of mediation analysis (Hox, 2013).

Conventional SEM/CFA software can estimate two-level models by treating the two levels as two groups (Muthén, 1994). Mehta and Neale (2005) described in detail how multilevel models can be incorporated in SEM/CFA. However, because using conventional SEM/CFA software requires complicated model specifications, recent versions of most SEM software packages (EQS, Bentler, 2005; LISREL, Joreskog & Sorbom, 1989, 1993; Mplus, Muthén & Muthén, 1998-2012; and Stata, StataCorp, 2015). Some extensions of this approach permit the use of categorical and ordinal data, incomplete data, and >2 levels (Hox, 2013; Kline, 2016). These new capabilities are summarized next. For more detailed applications of the MLV approach in the literature please refer to Dedrick and Greenbaum (2010, 2011); Dyer, Hanges, and Hall (2005); Kaplan and Kreisman (2000); and J. Little (2013), Byrne (2012), Heck & Tomas (2015) and Brown (2015). See a summary of the main advantages of MLV in **Table 3**.

3. Description of Multilevel Factor Analysis

One important feature of multilevel modeling is the flexibility to decide whether the effects of micro-level variables are fixed to be the same across macro-level research units (called a *fixed effect*), or are permitted to vary—called a *random effect* (Darlington & Hayes, 2017). Thus, *random coefficients* are parameters in a model that vary across clusters. Covariates could be included in a multilevel model to represent variability within and between clusters. To elaborate the example of classrooms nested within schools further, (see Brown, 2015 for a similar example), a multilevel regression model could examine, e.g., if a student's gender is a significant predictor of achievement in verbal ability. Gender would be a within-level effect (Level 1 or Micro level) because gender is a characteristic of individuals and the gender covariate illustrates variation in verbal achievement among individuals. An example of a between-level effect (Level 2), the age of the teacher (a classroom variable) may illustrate variability in oral achievement across classrooms. Thus, the effect of gender in oral achievement is a random slope (the slope varies across clusters) and the level 2 covariate of teacher age explained the variability of this coefficient across clusters/classrooms (Hox, 2010, 2013; Brown, 2015).

To return to the previous example of students nested within schools, this multilevel structure suggests that the total covariance matrix, Σ , would be divided into a within-covariance matrix Σ_W and a between-covariance matrix Σ_B . The Σ_W matrix contains covariances at the individual level (i.e., individual score differences in oral achievement) and their correlates accounting for variation across schools. In contrast, the Σ_B matrix represents covariation at the school level (i.e. differences across schools in the teaching experience and age of the teaching staff). The Σ_W and Σ_B covariance matrices can either have similar or totally different factor structures (Byrne, 2012). For each student in Level 1, the total score comprises a Level 1 component accounting for the individual deviation from the group mean and a Level 2 component accounting for the disaggregated school group mean. This individual composition allows separate calculation of within- and between-group covariance matrices (Heck, 2001; Hox, 2002 as quoted by Byrne, 2012). The related effects are defined *within-cluster effects* and *between-cluster effects* (Bentler, 2005). If a mean structure is necessary, it is used to illustrate the between-group means (Byrne, 2012).

In two-level structures, the observed individual-level variables are calculated by the following within and between equations:

$$y_W = \Lambda_W \eta_W + \varepsilon_W \quad (1) \text{ (within level)}$$

$$\mu_B = \mu + \Lambda_B \eta_B + \varepsilon_B \quad (2) \text{ (between level)}$$

μ = vector of between-level means

Λ_W = within-level factor loading matrix

Λ_B = between-level factor loading matrix

η_W = within-level factor

η_B = between-level factor

ε_W = within-level indicator residual variance

ε_B = between-level indicator residual variance

(Hox, 2013: p. 287; Brown, 2015: p. 421)

In the first equation the within-groups variation is represented. The second equation denotes the between-groups variation and the group level means while the factor loading matrices (Λ_W , Λ_B) and cluster-level means μ are considered *fixed effects* (Brown, 2015). Importantly, μ_B represents the random intercepts of the X variables that are the focus of the between-level means. By their combination Equation (3) is obtained:

$$X_{ij} = \mu + \Lambda_W \eta_W + \Lambda_B \eta_B + \varepsilon_B + \varepsilon_W \quad (3)$$

μ = vector of between-level means

Λ_W = within-level factor loading matrix

Λ_B = between-level factor loading matrix

η_W = within-level factor

η_B = between-level factor

ε_W = within-level indicator residual variance

ε_B = between-level indicator residual variance

(Hox, 2013; p. 288; Brown, 2015; p. 422)

Equation (2) is similar to equations used by random intercept regression models (except for symbols), with the loadings in the place of fixed regression coefficients and the factor matrices and a level-one and level-two error term. By allowing variation at the group-level factor loadings, this model is a generalized *random coefficient model*. The model in Equation (3) is a two-level factor model. If we add structural relationships between the latent factors at both levels, a multilevel SEM/CFA with two levels derives (Hox, 2013; Brown, 2015).

Multilevel models can be employed to analyze both EFA and CFA models. Actually, the within and between levels might have different number of latent variables, because, applied research suggests that typically fewer factors emerge at the between levels than at within levels because the variability across groups is lower than among individuals. Any CFA parameter (like factor loadings, or indicator intercepts) might be handled like a *random coefficient*, if justifiable by substantive theory and are based on empirical basis (Brown, 2015). Additionally, more complex data structures like cross-classifications, multiple-memberships or covariates are only few of the possible extensions of the basic CFA models developed (see Goldstein & Browne, 2005; Byrne, 2012).

Another feature of the multilevel CFA modeling is the disintegration of the total variance (Ψ) of the latent variables into the part attributed to between-cluster variation (Ψ_B) and the part attributed to within-cluster variation (Ψ_W). Based on these variances, the intraclass correlation (ICC) for the indicators can be estimated as:

$$ICC = \frac{\Psi_B}{\Psi_B + \Psi_W} \quad (4)$$

(Finch & Bolin, 2017; p. 237)

ICC values can range from 0.0 to 1.0 (Byrne, 2012). Generally, if the ICCs are all small, e.g., <0.05, the between-group variance is low and possibly there is no need to specify an MLV CFA model (Hox, 2013; Brown, 2015). Muthén (1997) noted—as reproduced by Byrne (2012)—that while ICC values usually range from 0.00 to 0.50 ICC values of 0.10 or larger, for a group size of 15 or larger suggest that MLV data should definitely be modeled. However, Julian (2001) and Selig et al. (2008) cautioned that even with ICC < 0.10, the hierarchical structure of the data should be taken into account (Byrne, 2012). Mehta and Neale (2005) proposed a method to compare the factor variances at levels 1 and 2. Specifically, (as reproduced by Finch & Bolin, 2017 and Heck & Thomas, 2015) the factor loadings across levels must be invariant. Thus, the loadings for each indicator at level 1 are constrained to be equivalent to the corresponding loading at level 2.

4. A Walk-Through into the Multilevel CFA

Multilevel CFA models are evaluated in multiple steps (Hox, 2013). Byrne (2012) states that three different methods emerged over the years. The first was a method proposed by Muthén (1994) initially containing four phases with the

MUML as an estimator. Muthén (1989, 1990, 1991, 1994) simplified the multi-level data analysis by using conventional SEM software by computing separate within and between-groups covariance matrices, which are orthogonal (uncorrelated) and additive (Heck & Thomas, 2015). However, as Byrne (2012) comments, elaboration of the MLV modeling estimation—moving from MUML to FIML—plus the evolution of statistical software used (Kaplan et al., 2009) and Bayesian methods of estimation (Heck & Thomas, 2015) inevitably altered the original methodology proposed by Muthén (1994). Specifically, Byrne (2012) explains that some phases (2 - 4) were unified (c.f. Mplus, 1998-2012). The second method was proposed by Hox (2002) and it tests the fundamental assumptions of MLV modeling by establishing benchmark models. Finally, the third method was developed by Mehta and Neale (2005) and is based on a process of 3 phases of fitting the univariate random intercepts to the data. The Hox (2002) method is described as the most uncomplicated to carry out (Selig et al., 2008; Byrne, 2012), but the Muthén (1994) approach is still the most frequently used (Cheung & Au, 2005; Byrne, 2012). See Byrne (2012) for details. A brief description of the steps of the most widely used method proposed by Muthén (1994) or the general-specific method (Heck & Thomas, 2015) follows.

The Steps of the method

The following steps were described by Hox (2013) for regression and SEM models and they were further detailed for CFA models by Brown (2015) and by Heck and Thomas (2015), Byrne (2012) and Finch and Bolin (2017). The following three steps are suggested for the estimation of a two-level model with the within-structure fully specified (Hox, 2013; Brown, 2015). This method was originally proposed by Muthén (1994) using MUML estimator. However, as Byrne (2012) comments, subsequent elaboration of MLV modeling—from MUML estimator to FIML estimator—plus evolution of statistical software simplified the method. A two-level model can be analyzed following three steps.

- **Step 1:** *The intraclass correlations* of the indicators are first examined (ICCs) of the indicators to examine group-level properties, i.e., how much variance in the indicator is explained by group membership (Shumacker & Lomax, 2016). In other words, to examine the extent of individual scores dependency within groups due to similarities of individuals (Field, 2013; Brown, 2015, Byrne, 2012; Tabachnick & Fidell, 2013; Kalaian & Kasim, 2007). The higher the ICC, the more score variance is attributed to the stratification or cluster (grouping variable). Using a design effect to estimate the difference between a multi-level nested design is possible as compared to a simple random sample (Shumacker & Lomax, 2016). As an alternative, using different non-hierarchical methods is possible, that do allow for a certain minor dependency in the data (Brown, 2015; Muthén & Muthén, 1998-2012). If the between-group variances are substantial, then the between structure is necessary to be taken into account (Hox, 2013).
- **Step 2:** Then the data of the within structure is analyzed (Level 1). At this

level (the individual level) a standard CFA is used (Hox, 2013) to ensure a viable measurement model at the within level with the between level unstructured (Brown, 2015), or (beyond CFA) more generally statistical techniques for clustered samples (cf. de Leeuw, Hox, & Dillman, 2008). First, we carry out a CFA to test the validity of the hypothesized structure based on the covariance matrix of the full sample, without taking into account the data hierarchy. If model modifications suggested by MIs are supported by substantive theory, the model can be re-specified accordingly to include additional parameters used for the individual level only (Byrne, 2012: p. 355). The fit of the model is then examined with conventional fit criteria (e.g., Hu & Bentler, 1999) and if satisfactory the researcher proceeds to the next step. As a rule, fit indices used are (Byrne, 2012): χ^2 , Comparative Fit Index (CFI; Bentler, 1990), Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980), and Standardized Root Mean Square Residual (SRMR).

- **Step 3:** if an acceptable measurement model emerges, the final step is to examine the between-level factor structure (Level 2) with the within-level factor structure (Level 1) completely modeled (Hox, 2013; Brown, 2015). Many MLV models with latent variables found in literature, but few of them are psychometrically oriented (Dedrick & Greenbaum, 2010; Byrne 2012). With an adequate fit for the single-level CFA model, then the factor structure of both individual and group level-data are tested simultaneously. Analyses can be based on robust Maximum likelihood (MLR; Muthén & Muthén, 1998-2012) estimator. However, in this step an error message may occur related to the higher level of the model. That is the higher level of the model must be overidentified for the model to be estimated properly. Specifically, error messages occur due to the usually small sample of the higher level. Unluckily, even when estimated parameters at the higher level are adequate (i.e., the model is over-identified), the same error message may again appear (c.f. Byrne, 2012). Note that by using the variance—covariance formula [$p(p + 1)/2$] estimating the number of variance—covariance parameters when a group level is added is possible. However, in MLV CFA the number of variance—covariance parameters doubled and the k intercept parameters estimated at Level-2 are added (Heck & Thomas, 2015). If presented with persistent error messages Byrne (2012) proposes to consider carrying out the MLV CFA analysis using the MUMML estimator instead of the MLR. Note however, that MUMML cannot handle deviations from multivariate normality. According to studies on the MUMML (Hox & Maas, 2001; Yuan & Hayashi, 2005) when using MUMML the likelihood of inadmissible solutions is greater if the sample size at the higher level is less than 50 (quoted in Byrne, 2012). In an admissible solution, according to Hox and Maas (2001) as reproduced by Byrne (2012), as a rule the factor loadings are generally accurate, but the residual variances and the standard errors may be underestimated.

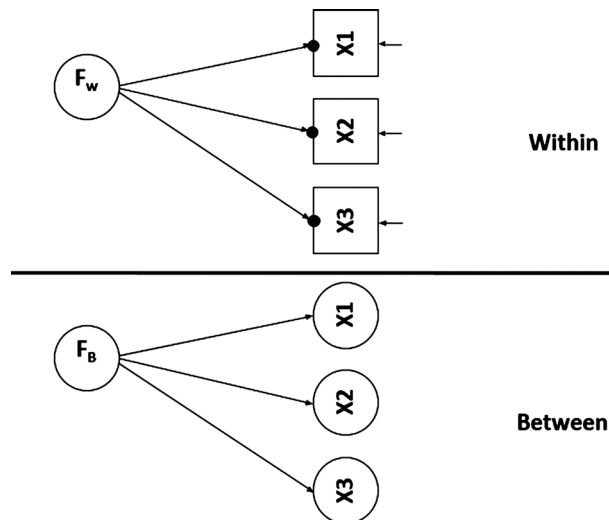
If the estimation of the model will produce no errors the initial information

examined is the following: (a) model summary results like the number of clusters in the analysis and the average cluster size, and (b) the ICCs pertinent to each of the observed variables. If the ICCs of the observed variables calculated in this step based on the simultaneous analysis at both levels are >0.10 (see Muthén, 1997 and Byrne, 2012), then the continuation of MLV analysis is supported (Julian, 2001; Byrne, 2012). Model Fit is evaluated by the following measures (Byrne, 2012): Chi-Square Test of Model Fit, Comparative Fit Index (CFI; Bentler, 1990), Tucker Lewis Index (TLI; Tucker & Lewis, 1973), Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980), and Standardized Root Mean Square Residual (SRMR) for the within model, Standardized Root Mean Square Residual (SRMR) for the between model (Byrne, 2012). Akaike's Information Criterion (AIC; Akaike, 1987), and the Bayesian Information Index (BIC; Raftery, 1993; Schwartz, 1978) can also be used for MLV model fit comparison (Heck & Thomas, 2015). Crucially, even if model fit is acceptable, the estimated parameters must be examined as well to decide if the model is acceptable (i.e. significant factor loadings and relatively low measurement errors). These goodness-of-fit indices apply to the entire model. Specifically, they show to what extent the model fits the within-group model data and of the between-group model. Moreover, the likelihood function can be used for the calculation of the deviance statistic by multiplying with -2 ($-2LL$ log likelihood function²), where the log is the natural logarithm and likelihood is the value of the likelihood function at convergence (Heck & Thomas, 2015). Generally, models with lower deviance show better fit than models with higher deviance (Hox, 2002; Heck & Thomas, 2015).

Consider an example (Figure 3) where a 3-item questionnaire assessing Parental Satisfaction, e.g. Kansas Parental Satisfaction Scale (James et al., 1985) is administered to 620 parents from 31 different neighborhoods. This is a two-level data structure where parents are nested under neighborhoods with an average cluster size of 20. Note that the size of the sample at the individual level (Level 1) of the hierarchical structure is 620 and that at the group level (Level 2) is 31. The study objective would be to create a multilevel CFA model with two levels to take into consideration the variability of a hypothesized 3-item single-factor questionnaire (see Figure 3) with continuous data given the existing suggestions that as the number of scale points increases, ordinal data (>5 or 7 Likert points) then interval data are treated like continuous (Boomsma, 1987; Rigdon, 1998; Byrne, 2012).

The path diagram in Figure 3 follows the conventions proposed by Muthén and Muthén (1998-2012). Note that indicators ($x_1 - x_3$) are within-levels observed variables (parents), but at the between level (neighborhoods) they become latent variables. Therefore, the black circles of the 3 indicators shown in the path diagram are continuous random intercepts for the observed items that vary across clusters. Observed variables for each individual are assumed to have a unique, person specific, within-cluster source of variance (Mehta & Neale, 2005;

²See also Satorra & Bentler (2010).



Adapted by Mplus user Guide (Muthén & Muthén, 1998-2012)

Figure 3. MLV CFA model with one factor and two levels. In the between-groups model, the random intercepts are continuous latent variables. Thus, they are represented small black ovals at the end of the latent factors arrows to illustrate random intercepts (Muthén & Muthén, 1998-2012).

Heck & Thomas, 2015). On the between levels, the single factor (F_B) is specified to account for the variation and covariation among these random intercepts (Brown, 2015). For a similar applied example, the readers can refer to Brown (2015). For instructions on how to extend the CFA model to three levels readers can refer to Heck & Thomas (2015).

Brown (2015) notes that the following parameters are freely assessed (Muthén & Muthén, 1998-2012): factor variances at both levels, fixed intercepts at the between-level and indicator residual variances at both levels (Brown, 2015). By default, the latent-variable means and covariances of the residuals are fixed to zero at both levels. Note that the magnitudes of the variances of the parental satisfaction factors at both levels are not directly comparable unless the factors have a common metric. If the within and between levels have the same measurement model, the equality of factor loadings across levels can be tested. The metrics of the within-and between-level factors will be equated if the factor loadings are equivalent. Therefore, factor variances will also be directly comparable (Mehta & Neale, 2005; Brown, 2015). However, if there is no common scale of measurement across levels, the magnitude of the factor variances at each level is not directly comparable (Mehta & Neale, 2005; Heck & Thomas, 2015). Consequently, establishing a common scale of measurement across levels is often useful (Heck & Thomas, 2015). Alternatively, Byrne (2012) follows the same procedure described above by omitting the initial calculation of the ICC. A second differentiation of the applied example proposed by Byrne (2012) is the inclusion of a different measurement model across levels. Finally, Byrne comments that ideally, to get a more accurate result description the model fit must be evaluated separately for each of the two levels. This procedure is described

next, along with other important issues in the MLV CFA.

5. Important Considerations of MLV CFA

Model Estimation

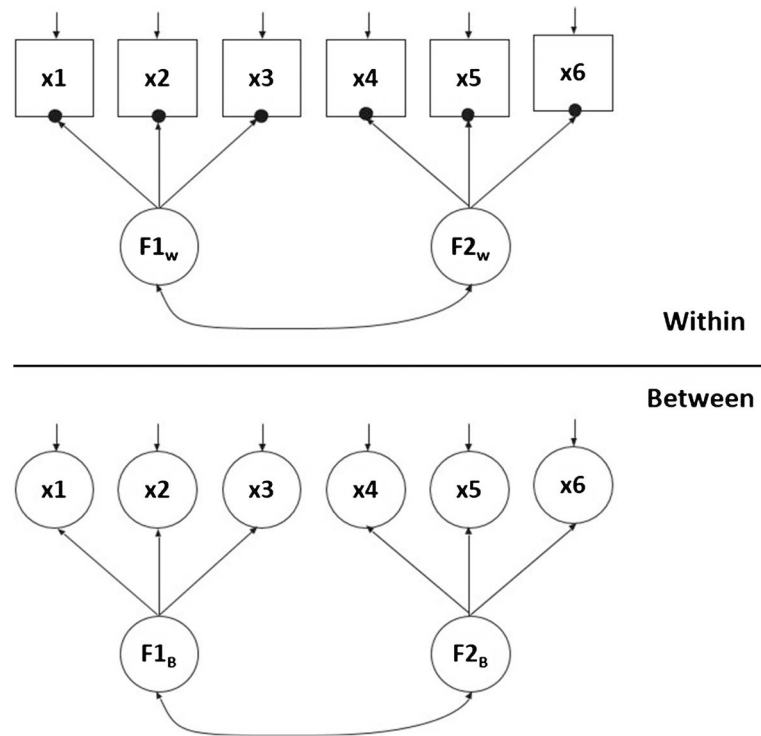
During early MLV SEM modeling—as [Byrne \(2012\)](#) describes—the parameter estimation was carried out mainly by full information maximum likelihood estimation (FIML) adjusted for multilevel data (MUML), and it was proposed by [Muthén \(1994\)](#). More recent advancements in SEM research brought about important refinements in ML estimation and MLV modeling ([Heck & Thomas, 2009](#); [Kaplan et al., 2009](#)). These newer estimation methods can be distinguished based on their approach to the computation of standard errors. The first of these methods is based on the MLF estimator; the second is based on the usual ML estimator on second-order derivatives and the third is based on the MLR estimator, which is robust to nonnormality but also permits MLV analyses based on unbalanced groups. Given these new possibilities it was suggested that MUML estimator may no longer be needed ([Yuan & Hayashi, 2005](#); [Byrne, 2012](#)). Obviously, these estimation options increased SEM MLV modeling flexibility adding computational power ([Heck & Thomas, 2009](#)). However, [Byrne \(2012\)](#) showed that MUML could be useful in case of errors generated during model estimation, typically caused by small sample size at levels > 1 .

Model fit evaluation

As [Finch and Bolin \(2017\)](#) argue, fit statistics—maybe except Standardized Root Mean Square Residual (e.g. in Mplus; [Muthén & Muthén, 1998-2012](#))—typically present combined model fit information about both levels (also [Byrne, 2012](#)). Usually points at Level 1 are greater than those of Level 2, fit indices primarily measure the level 1 model fit ([Ryu & West, 2009](#); [Byrne, 2012](#)). [Stapleton \(2013\)](#) provides instructions on separate model fit evaluation at each level, reproduced here based on [Finch and Bolin \(2017\)](#).

First, the Chi-square model fit statistics are calculated for Level 1 baseline models. This process is repeated for Level 2 baseline models. To calculate the baseline value for the level 1 part of the model, the covariances of the observed indicators are constrained to 0 at Level 1 and they are freely estimated at Level 2. By this method, the baseline Chi-square fit statistic is obtained. Similarly, to estimate the baseline value for the level 2 part of the model, the covariances of the observed indicators are constrained to 0 at Level 2 and they are freely estimated at Level 1. See an example of the path diagram of an MLV CFA Model with two factors in [Figure 4](#).

Following [Stapleton's \(2013\)](#) steps for calculating fit statistics at each level separately, a saturated model specified at level 2 (i.e. with a perfect fit at that level), with the level 1 model fully specified. The resulting Chi-square fit statistic is then examined. Using an equation described by [Ryu and West \(2009\)](#), the comparative fit index (CFI) for the level 1 part of the model is then obtained. Likewise, we can obtain the level 2 CFI value in a comparable method, i.e. by obtaining



Adapted by Mplus user Guide (Muthén & Muthén, 1998-2012)

Figure 4. MLV CFA model with two factors and two levels. Again in the between—groups model, the random intercepts are continuous latent variables, represented by small black ovals to model random intercepts (Muthén & Muthén, 1998-2012).

the Chi-square goodness of fit statistic for the level 1 saturated model. The fit at level 2 is then examined following typical CFA guidelines (e.g. Hu & Bentler, 1999; Brown, 2015; Kline, 2016). Based on previous analyses the CFI values at levels 1 and 2 are both examined. The SRMR is also examined providing model fit information at both levels separately. From these results, we can decide if model fit at each level is acceptable. If the model shows a good fit to the data, the evaluation of model parameters comes next.

Moreover, to estimate the amount of variance of the observed indicators, attributed to each data level, the same latent structure at each level must be specified and the factor loadings must be constrained to equality at both levels (Mehta & Neale, 2005; Brown, 2015; Finch & Bolin, 2017; Heck & Thomas, 2015). In order to obtain the ICC for each factor, we would employ Equation 4 above, using the factor variances resulting from this constrained model (Finch & Bolin, 2017).

Sample Size

ML estimation is notorious for requiring large sample sizes, and this is also true for MLV CFA (Heck, 2001; Hox, 2002; Hox & Maas, 2001; Muthén, 1994; Yuan & Bentler, 2002, 2004; Byrne, 2012). As a rule, in multilevel modeling, the sample size of the highest level is generally of primary importance, because the higher level sample sizes are smaller than the lower level sample sizes (Hox,

2013). A minimum sample size of 60 for the highest level was recommended by Eliason (1993) when using ML as an estimator (Hox, 2013). However, Maas and Hox (2005) set this value to at least 100 groups, although for uncomplicated models, even 50 groups may also suffice. Although these recommendations were initially made for regression models, multilevel regression accuracy of higher level variances also applies to SEM and CFA models, because multilevel SEM is also based on the within-group and between-group covariance matrices (Hox, 2013; Hox, Maas, & Brinkhuis, 2010).

Finally, an unequal sample size at each level is not a problem for the estimation of the model, because unequal sample sizes are assumed by FIML (MLR or MLM). However, the interpretation of model fit indicators must be made cautiously. Also in longitudinal models, missing values commonly attributed to missing occasions or panel dropout can be easily handled. Van Buuren (2011) has elaborated on incomplete multilevel data (Hox, 2013). For in depth analysis of the theoretical background and applied examples of MLV modeling the following resources are referred: Bovaird (2007), Heck (2001), Hox (2002, 2010), Kaplan et al. (2009), Little et al. (2000), Reise and Duan (2003), Selig et al. (2008), Hoffman (2007), Byrne (2012), Heck & Thomas (2015) and Finch & Bolin (2017). Typically, when MLV CFA is carried out to establish construct validity additional analyses are required (Byrne, 2012), that are beyond the scope of this work. However, a detailed description a multi-phased method of construct validity was provided by Kyriazos (2018) or applied examples by Kyriazos, Stalikas, Prassa, Yotsidi (2018a, 2018b), Kyriazos, Stalikas, Prassa, Yotsidi, Galanakis, Pezirkianidis (2018) and Kyriazos, Stalikas, Prassa, Galanakis, Yotsidi, Lakioti (2018). Some specialized applications of MLV modeling within the SEM framework are recommended by SEM experts (Byrne, 2012). Specifically, for information on longitudinal analyses and/or latent growth curve modeling, Byrne refers readers to Chen, Kwok, Luo, and Willson (2010); Chou, Bentler, and Pentz (1998); Ecob and Der (2003); Hung (2010); Jo and Muthén (2003); Kwok, West, and Green (2007); MacCallum and Kim (2000); and Muthén, Khoo, Francis, and Boscardin (2003).

6. Summary & Conclusion

Multilevel modeling is an extremely complicated topic. We can only skim the surface of these intriguing sets of methods (Tabachnick & Fidell, 2013; Field 2013). The need for multilevel modeling arose because data are sometimes collected from people or other units that are “nested” in some fashion under different higher-level research units (Darlington & Hayes, 2017). These hierarchical models explicitly model lower and higher levels by taking into account the interdependence of individuals within each sample group. In MLV CFA analysis, biases in parameter estimates, standard errors, and tests of model fit emerge if the hierarchical structure is ignored, additionally, if the nonindependence of the observations and the standard errors of parameter estimates may be underesti-

mated, resulting in positively biased statistical significance testing (Brown, 2015).

The MLV CFA procedure assumed that latent factors contain between- and within-group elements. The between-group element is typically the general part and the within-group element is the specific part of the model. During this process, if factor loadings are constrained to be invariant with its counterpart, factor loadings on the different level provide a way to equate factor scales across levels thus, enabling the direct comparison of factor variances across levels (Mehta & Neale, 2005; Heck & Thomas, 2015; Brown, 2015; Finch & Bolin, 2017).

Alternatively, a different method can be used for evaluating clustered data and this is the hierarchical factor model (Bauer, 2003; Curran, 2003; Harnqvist, Gustafsson, Muthén, & Nelson, 1994; Mehta & Neale, 2005; cited in Heck & Thomas, 2015). During this procedure, the assumption of invariant factor loadings across levels is examined but additionally, the assumption of zero variability of the observed indicators at the cluster level is also evaluated (Mehta & Neale, 2005; Heck & Thomas, 2015). If these two assumptions are true, a hierarchical factor model emerges, where latent variables at the individual level define the latent factor at the higher level (Mehta & Neale, 2005; Heck & Thomas, 2015).

During the estimation of multilevel CFA models, problems frequently arise. Specifically, due to the smaller sample size, or the necessity of a more parsimonious structure, the between-group structure may be more prone to errors during model estimation. Researchers are also in debate whether missing data are a problem (Heck & Thomas, 2015) or not (Hox, 2013). The MLV CFA is carried out in several steps. If a problem occurs, defining model starting values to boost the iteration of the software to a solution may be necessary (Heck & Thomas, 2015). Alternatively, it could help to specify the model progressively, e.g. by defining one factor at a time within the multilevel model. Sometimes this could help to learn exactly where the problem lies. Anyhow, patience is a valuable strength of character when carrying out a multilevel CFA (Heck & Thomas, 2015)!

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Ainsworth, M. D. S., Blehar, M. C., & Waters, E. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation*. Hillsdale, NJ: Earlbaum.
- Akaike, H. (1987). Factor Analysis and AIC. *Psychometrika*, *52*, 317-332. <https://doi.org/10.1007/BF02294359>
- Bauer, D. J. (2003). Estimating Multilevel Linear Models as Structural Models. *Journal of Educational and Behavioral Statistics*, *28*, 135-167. <https://doi.org/10.3102/10769986028002135>
- Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, *107*, 238-246. <https://doi.org/10.1037/0033-2909.107.2.238>

- Bentler, P. M. (2005). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multi-variate Software.
- Bickel, R. (2007). *Multilevel Analysis for Applied Research: It's Just Regression!* New York: Guilford.
- Boomsma, A. (1987). The Robustness of Maximum Likelihood Estimation in Structural Equation Modeling. In P. Cuttance, & R. Ecob (Eds.), *Structural Equation Modelling by Example* (pp. 160-188). Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511601118.010>
- Bovaird, J. A. (2007). Multilevel Structural Equation Models for Contextual Factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling Contextual Effects in Longitudinal Studies* (pp. 149-182). Mahwah, NJ: Erlbaum.
- Bowlby, J. (1969). *Attachment and Loss: Attachment* (Vol. 1). London: Penguin.
- Bowlby, J. (1973). *Attachment and Loss Vol. 2. Separation: Anxiety and Anger*. New York: Basic Books.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York: Guilford Publications.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of Hierarchical Linear Models to Assessing Change. *Psychological Bulletin*, *101*, 147-158.
<https://doi.org/10.1037/0033-2909.101.1.147>
- Byrne, B. M. (2012). *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. London: Routledge. <https://doi.org/10.4324/9780203807644>
- Chen, Q., Kwok, O.-M., Luo, W., & Willson, V. L. (2010). The Impact of Ignoring a Level of Nesting Structure in Multilevel Growth Mixture Models: A Monte Carlo Study. *Structural Equation Modeling*, *17*, 570-589.
<https://doi.org/10.1080/10705511.2010.510046>
- Cheung, M. W.-L., & Au, K. (2005). Applications of Multilevel Structural Equation Modeling to Cross-Cultural Research. *Structural Equation Modeling*, *12*, 598-619.
https://doi.org/10.1207/s15328007sem1204_5
- Chou, C.-P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of Two Statistical Approaches to Study Growth Curves: The Multilevel Model and the Latent Curve Analysis. *Structural Equation Modeling*, *5*, 247-266.
<https://doi.org/10.1080/10705519809540104>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Erlbaum.
- Curran, P. J. (2003). Have Multilevel Models Been Structural Equation Models All Along? *Multivariate Behavioral Research*, *38*, 529-569.
https://doi.org/10.1207/s15327906mbr3804_5
- Darlington, R. B., & Hayes, A. F. (2017). *Regression Analysis and Linear Models: Concepts, Applications, and Implementation*. New York: The Guilford Press.
- de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). *The International Handbook of Survey Methodology*. New York: Taylor & Francis.
- Dedrick, R. F., & Greenbaum, P. E. (2010). Multilevel Confirmatory Factor Analysis of a Scale Measuring Interagency Collaboration of Children's Mental Health Agencies. *Journal of Emotional and Behavioural Disorders*, *10*, 1-14.
<https://doi.org/10.1177/1063426610365879>
- Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel Confirmatory Factor Analysis of a Scale Measuring Interagency Collaboration of Children's Mental Health Agencies. *Journal of Emotional and Behavioral Disorders*, *19*, 27-40.

- <https://doi.org/10.1177/1063426610365879>
- Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying Multilevel Confirmatory Factor Analysis Techniques to the Study of Leadership. *Leadership Quarterly, 16*, 149-167. <https://doi.org/10.1016/j.leaqua.2004.09.009>
- Ecob, R., & Der, G. (2003). An Iterative Method for the Detection of Outliers in Longitudinal Growth Data Using Multilevel Models. In S. P. Reise, & N. Duan (Eds.), *Multilevel Modeling: Methodological Advances, Issues, and Applications* (pp. 229-254). Mahwah, NJ: Erlbaum.
- Eliason, S. R. (1993). *Maximum Likelihood Estimation*. Newbury Park, CA: Sage. <https://doi.org/10.4135/9781412984928>
- Epstein, N. B., Bishop, D. S., & Levin, S. (1978). The McMaster Model of Family Functioning. *Journal of Marital and Family Therapy, 4*, 19-31. <https://doi.org/10.1111/j.1752-0606.1978.tb00537.x>
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). London: Sage.
- Finch, W. H., & Bolin, J. E. (2017). *Multilevel Modeling Using Mplus*. Boca Raton, FL: Taylor & Francis Group. <https://doi.org/10.1201/9781315165882>
- Geiser, C. (2013). *Data Analysis with Mplus* (English Edition). New York: The Guilford Press.
- Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. London: Oxford University Press.
- Goldstein, H., & Browne, W. (2005). Multilevel Factor Analysis Models for Continuous and Discrete Data. In A. Maydeu-Olivares, & J. J. McArdle (Eds.), *Contemporary Psychometrics: A Festschrift to Roderick P. McDonald* (pp. 453-475). Mahwah, NJ: Lawrence Erlbaum Associates.
- Harnqvist, K., Gustafsson, J., Muthén, B., & Nelson, G. (1994). Hierarchical Models of Ability at Class and Individual Levels. *Intelligence, 18*, 165-118. [https://doi.org/10.1016/0160-2896\(94\)90026-4](https://doi.org/10.1016/0160-2896(94)90026-4)
- Heck, R. H. (2001). Multilevel Modeling with SEM. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 89-127). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Heck, R. H., & Thomas, S. L. (2009). *An Introduction to Multilevel Modeling Techniques* (2nd ed.). New York: Routledge.
- Heck, R. H., & Thomas, S. L. (2015). *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus* (3rd ed.). New York: Routledge. <https://doi.org/10.4324/9781315746494>
- Hoffman, L. (2007). Multilevel Models for Examining Individual Differences in Within-Person Variation and Covariation over Time. *Multivariate Behavioral Research, 42*, 609-629. <https://doi.org/10.1080/00273170701710072>
- Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410604118>
- Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). New York: Routledge. <https://doi.org/10.4324/9780203852279>
- Hox, J. J. (2013). Multilevel Regression and Multilevel Structural Equation Modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (pp. 281-294). New York: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199934898.013.0014>
- Hox, J. J., & Maas, C. J. M. (2001). The Accuracy of Multilevel Structural Equation Mod-

- eling with Psuedobalanced Groups and Small Samples. *Structural Equation Modeling*, 8, 157-174. https://doi.org/10.1207/S15328007SEM0802_1
- Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The Effect of Estimation Method and Sample Size in Multilevel Structural Equation Modeling. *Statistica Neerlandica*, 64, 157-170. <https://doi.org/10.1111/j.1467-9574.2009.00445.x>
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling*, 6, 1-55. <https://doi.org/10.1080/10705519909540118>
- Hung, L.-F. (2010). The Multigroup Multilevel Categorical Latent Growth Curve Models. *Multivariate Behavioral Research*, 45, 359-392. <https://doi.org/10.1080/00273171003680336>
- James, D. E., Schumm, W. R., Kennedy, C. E., Grigsby, C. C., Sheckman, K. L., & Nichols, C. W. (1985). Characteristics of the Kansas Parental Satisfaction Scale among Two Samples of Married Parents. *Psychological Reports*, 57, 163-169. <https://doi.org/10.2466/pr0.1985.57.1.163>
- Jo, B., & Muthén, B. O. (2003). Longitudinal Studies with Intervention and Noncompliance: Estimation of Causal Effects in Growth Mixture Modeling. In S. P. Reise, & N. Duan (Eds.), *Multilevel Modeling: Methodological Advances, Issues, and Applications* (pp. 112-139). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7: User's Reference Guide*. Chicago, IL: Scientific Software.
- Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago, IL: Scientific Software International.
- Julian, M. W. (2001). The Consequences of Ignoring Multilevel Data Structures in Non-hierarchical Covariance Modeling. *Structural Equation Modeling*, 8, 325-352. https://doi.org/10.1207/S15328007SEM0803_1
- Kalaian, S. A., & Kasim, R. M. (2007). Hierarchical Linear Modeling. In N. J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 432-435). Thousand Oaks, CA: Sage.
- Kaplan, D., & Kreisman, M. B. (2000). On the Validation of Indicators of Mathematics Education Using TIMSS: An Application of Multilevel Covariance Structure Modeling. *International Journal of Educational Policy, Research, and Practice*, 1, 217-242.
- Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel Latent Variable Modeling: Current Research and Recent Developments. In R. E. Millsap, & A. Maydeu-Olivares (Eds.), *The Sage Handbook of Quantitative Methods in Psychology* (pp. 592-613). Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9780857020994.n24>
- Kelloway, E. K. (2015). *Using Mplus for Structural Equation Modeling*. Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781849209366>
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.).
- Kreft, I., & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. Newbury Park, CA: Sage.
- Kwok, O.-M., West, S. G., & Green, S. B. (2007). The Impact of Misspecifying the Within-Subject Covariance Structure in Multiwave Longitudinal Multilevel Models: A Monte Carlo Study. *Multivariate Behavioral Research*, 42, 557-592. <https://doi.org/10.1080/00273170701540537>
- Kyriazos, T. A. (2018). Applied Psychometrics: The 3-Faced Construct Validation Method, a Routine for Evaluating a Factor Structure. *Psychology*, 9, 2044-2072. <https://doi.org/10.4236/psych.2018.98117>

- Kyriazos, T. A., Stalikas, A., Prassa, K., & Yotsidi, V. (2018a). Can the Depression Anxiety Stress Scales Short Be Shorter? Factor Structure and Measurement Invariance of DASS-21 and DASS-9 in a Greek, Non-Clinical Sample. *Psychology, 9*, 1095-1127. <https://doi.org/10.4236/psych.2018.95069>
- Kyriazos, T. A., Stalikas, A., Prassa, K., & Yotsidi, V. (2018b). A 3-Faced Construct Validation and a Bifactor Subjective Well-Being Model Using the Scale of Positive and Negative Experience, Greek Version. *Psychology, 9*, 1143-1175. <https://doi.org/10.4236/psych.2018.95071>
- Kyriazos, T. A., Stalikas, A., Prassa, K., Galanakis, M., Yotsidi, V., & Lakioti, A. (2018). Psychometric Evidence of the Brief Resilience Scale (BRS) and Modeling Distinctiveness of Resilience from Depression and Stress. *Psychology, 9*, 1828-1857. <https://doi.org/10.4236/psych.2018.97107>
- Kyriazos, T. A., Stalikas, A., Prassa, K., Yotsidi, V., Galanakis, M., & Pezirkianidis, C. (2018). Validation of the Flourishing Scale (FS), Greek Version and Evaluation of Two Well-Being Models. *Psychology, 9*, 1789-1813. <https://doi.org/10.4236/psych.2018.97105>
- Little, J. (2013). Multilevel Confirmatory Ordinal Factor Analysis of the Life Skills Profile-16. *Psychological Assessment, 25*, 810-825. <https://doi.org/10.1037/a0032574>
- Little, T. D., Schnabel, K. U., & Baumert, J. (2000). *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Scientific Examples*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410601940>
- Loehlin, J. C., & Beaujean, A. A. (2017). *Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis*. New York, NY: Taylor & Francis.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*, 85-91. <https://doi.org/10.1027/1614-2241.1.3.85>
- MacCallum, R. C., & Kim, C. (2000). Modeling Multivariate Change. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Scientific Examples* (pp. 51-68). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- McArdle, J. J., & Hamagami, F. (1996). Multilevel Models from a Multiple Group Structural Equation Perspective. In G. A. Marcoulides, & R. E. Schumaker (Eds.), *Advanced Structural Equation Modeling* (pp. 89-124). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Mehta, P. D., & Neale, M. C. (2005). People Are Variables Too: Multilevel Structural Equation Modeling. *Psychological Methods, 10*, 259-284. <https://doi.org/10.1037/1082-989X.10.3.259>
- Muthén, B. (2004). Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 345-368). Newbury Park, CA: Sage.
- Muthén, B. O. (1989). Latent Variable Modeling in Heterogeneous Populations. *Psychometrika, 54*, 557-585. <https://doi.org/10.1007/BF02296397>
- Muthén, B. O. (1990). Mean and Covariance Structure Analysis of Hierarchical Data. *Paper Presented at the Psychometric Society Meeting*, Princeton, June 1990, UCLA Statistics Series 62.
- Muthén, B. O. (1991). Multilevel Factor Analysis of Class and Student Achievement Components. *Journal of Educational Measurement, 28*, 338-354. <https://doi.org/10.1111/j.1745-3984.1991.tb00363.x>

- Muthén, B. O. (1994). Multilevel Covariance Structure Analysis. *Sociological Methods and Research*, 22, 376-398. <https://doi.org/10.1177/0049124194022003006>
- Muthén, B. O. (1997). Latent Variable Modeling of Longitudinal and Multilevel Data. In A. E. Raftery (Ed.), *Sociological Methodology 1997* (pp. 453-481). Washington DC: American Sociological Association. <https://doi.org/10.1111/1467-9531.271034>
- Muthén, B. O., & Satorra, A. (1995). Complex Sample Data in Structural Equation Modeling. In P. Marsden (Ed.), *Sociological Methodology 1995* (pp. 216-316). Boston, MA: Blackwell. <https://doi.org/10.2307/271070>
- Muthén, B., Khoo, S.-T., Francis, D. J., & Boscardin, C. K. (2003). Analysis of Reading Skills Development from Kindergarten through First Grade: An Application of Growth Mixture Modeling to Sequential Processes. In S. P. Reise, & N. Duan (Eds.), *Multilevel Modeling: Methodological Advances, Issues, and Applications* (pp. 71-89). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (Seventh ed.). Los Angeles, CA: Muthén & Muthén.
- Nezlek, J. B. (2011). *Multilevel Modeling for Social and Personality Psychology*. London: Sage. <https://doi.org/10.4135/9781446287996>
- Raftery, A. E. (1993). Bayesian Model Selection in Structural Equation Models. In K. A. Bollen, & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 163-180). Newbury Park, CA: Sage.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Newbury Park, CA: Sage.
- Reise, S. P., & Duan, N. (2003). *Multilevel Modeling: Methodological Advances, Issues, and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410606747>
- Rigdon, E. (1998). Structural Equation Models. In G. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 251-294). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Ryu, E., & West, S. G. (2009). Level-Specific Evaluation of Model Fit in Multilevel Structural Equation Modeling. *Structural Equation Modeling*, 16, 583-601. <https://doi.org/10.1080/10705510903203466>
- Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-Square Test Statistic. *Psychometrika*, 75, 243. <https://doi.org/10.1007/s11336-009-9135-y>
- Schumacker, R. E., & Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling* (4th ed.). New York: Routledge.
- Schwartz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464. <https://doi.org/10.1214/aos/1176344136>
- Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent Variable Structural Equation Modeling in Cross-Cultural Research: Multigroup and Multilevel Approaches. In F. J. R. van de Vijver, D. A. van Hemert, & Y. H. Poortinga (Eds.), *Multilevel Analysis of Individuals and Cultures* (pp. 93-119). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195152968.001.0001>
- Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Newbury Park, CA: Sage.
- Stapleton, L. M. (2013). Incorporating Sampling Weights into Single- and Multi-Level

- Models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *A Handbook of International Large-Scale Assessment* (pp. 353-388). London: Chapman Hall/CRC Press.
- StataCorp (2015). *Stata: Release 14. Statistical Software*. College Station, TX: StataCorp.
- Steiger, J. H., & Lind, J. C. (1980). Statistically Based Tests for the Number of Common Factors. *Paper Presented at the Psychometric Society Annual Meeting*, Iowa City, IA.
- Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics*. Boston, MA: Pearson Education Inc.
- Tucker, L. R., & Lewis, C. (1973). A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika*, *38*, 1-10. <https://doi.org/10.1007/BF02291170>
- van Buuren, S. (2011). Multiple Imputation of Multilevel Data. In J. J. Hox, & J. K. Roberts (Eds.), *Handbook of Advanced Multilevel Analysis* (pp. 173-196). New York: Routledge.
- Yuan, K.-H., & Bentler, P. M. (2002). On Normal Theory Based Inference for Multilevel Models with Distributional Violations. *Psychometrika*, *67*, 539-562. <https://doi.org/10.1007/BF02295130>
- Yuan, K.-H., & Bentler, P. M. (2004). On the Asymptotic Distributions of Two Statistics for Two-Level Covariance Structure Models within the Class of Elliptical Distributions. *Psychometrika*, *69*, 437-457. <https://doi.org/10.1007/BF02295645>
- Yuan, K.-H., & Hayashi, K. (2005). On Muthen's Maximum Likelihood for Two-Level Covariance Structure Models. *Psychometrika*, *70*, 147-167. <https://doi.org/10.1007/s11336-003-1070-8>