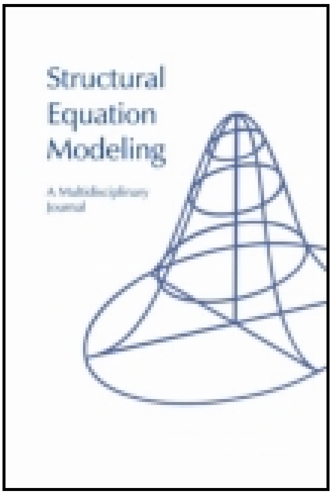

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd Edition)

Robert S. Stawski [a]
[a] University of Michigan
Published online: 22 Jul 2013.

Structural Equation Modeling, 20:541–550, 2013

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2013.797841

BOOK REVIEW

Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd Edition), Tom A. B. Snijders & Roel J. Bosker. Los Angeles, CA: Sage, 2012, 354 pp. $45.00 (softcover).

Reviewed by Robert S. Stawski
University of Michigan

In the second edition of *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling,* Snijders and Bosker have provided a detailed summary of the conceptual and practical application of multilevel models for use with data including design-based dependencies (e.g., nested data). The past decades in social and behavioral sciences have seen a dramatic increase in the number of studies incorporating complex sampling and design characteristics, making the new edition of this book an excellent resource and timely addition to the available literature and a welcome expansion of the previous edition of this book.

BOOK SUMMARY

This second edition of *Multilevel Analysis* consists of 18 chapters. The book starts with prefaces to the second edition, as well as the first edition, with the former providing highlights of the major additions and changes to the new offering (i.e., new chapters, revisions to existing chapters, revised approaches to empirical examples). The rationale for the structure of the book contained in the preface, focusing on "basic" (e.g., foundations of nested data/data with dependencies, variability at difference levels of analysis, random effects models) and "advanced" (e.g., Bayesian estimating, survey sampling/sampling weights, latent class models), will help provide readers with a frame of reference to maximize the information for their needs.

Chapter 1 provides a brief overview of the concept of multilevel analysis and the notion of complex sources of variability that exist with a focus on nested data (e.g., pupils in classes, longitudinal measurements of research participants, etc.). Here, the reader is introduced to the importance of considering variability at different levels of analysis, drawing on seminal work from Robinson (1950) on the ecological fallacy and the inherent problem of confusing aggregate and individual-level effects in nested or multilevel data structures (see also Borsboom, Mellenbergh, &Van Heerden, 2003; Molenaar, 2004). This conceptual point is used as the springboard to formally introduce the reader to what mixed models are and the history behind their (necessary) development. The remainder of Chapter 1 is devoted to a brief overview of the subsequent chapters in the book and details regarding the Web site supporting this book where data sets and scripts for replicating examples can be found. It also provides a brief detailing of the notation that is used for all equations presented in the book, which is helpful for those not as familiar with the Greek alphabet, and helps lay the foundation for conceptual continuity with respect to the more formal mathematical treatments throughout.

Chapter 2 provides a treatment of the theory behind multilevel data structures, the concept of multistage sampling (and sampling with dependencies more generally), and the importance of using appropriate methods for drawing inferences when analyzing data with complex multilevel structures. The majority of this chapter is devoted to introducing the reader to different situations where multilevel sampling strategies can be used and when multilevel models should be employed. Concepts such as *macrolevel* (e.g., classrooms) and *microlevel* (e.g., students within classrooms) units and the sampling of units that are related in some logical way are also conveyed with tables and figures, which help to crystallize these concepts. Furthermore, the concepts of dependence in the data as a nuisance, threat to the assumption of more traditional statistical models, and as an interesting phenomenon, and an opportunity to examine sources of systematic variation across different levels of study design features are addressed. Much of the chapter is devoted to nested data involving classical examples from education (e.g., students within classrooms), but it also introduces longitudinal or repeated assessments of individuals as a special case of nesting, which helps set the stage for readers to better understand complementarity between repeated measures data and nested data and how the same analytic approach applies to both cases. Finally, Chapter 2 introduces concepts regarding theoretical and statistical prediction involving *macrolevel* units, *microlevel* units, and cross-level interactions (e.g., moderation of the microlevel predictor by a macrolevel predictor). This section comes fully equipped with figures graphically depicting the multitude of opportunities that multilevel data and analysis can afford, but also the considerations that such data and analysis requires.

Chapter 3 provides a very thorough treatment of statistical considerations of multilevel data. Substantial coverage is given to the importance of considering all levels in one's multilevel data, and how failure to do so can result in erroneous parameter estimates and conclusions (e.g., the ecological fallacy; Robinson, 1950). Ways of considering the use of multilevel data including aggregation (e.g., collapsing/averaging lower levels of data so they are reflected by a single estimate at a higher level) and disaggregation (e.g., failing to acknowledge the nesting structure and treating the lowest level observations as independent) are presented to set the stage for introducing readers to the examination and analysis of variability at different levels of analysis. Here, readers are introduced to the intraclass correlation coefficient (ICC) for decomposing variability in multilevel data, and how variability at different levels of analysis warrant more nuanced considerations regarding the sampling and sample size of observations at different levels in a multiple-stage sampling design. This information serves as the backdrop for the first foray into the analysis of multilevel data through examples of between-group (Level 2) and within-group (Level 1) regression and correlation.

Chapter 4 covers the basic nuts and bolts of the multilevel model via treatment of the random intercept model. Here, readers are introduced to the concepts of fixed effects (i.e., sample average parameters) and random effects (i.e., variation in fixed effects across Level 2 units), and when random effects and random coefficient models should be employed. The empirical examples in this chapter (modeling variability and predicting language scores in elementary school—which are built on throughout the book) walk readers through the equations, analysis, and results of an empty or unconditional multilevel model, with considerable attention to the precision of interpretation of estimates. Following the treatment of the empty model, the authors introduce a simple extension via the inclusion of a single fixed effect predictor (student IQ) to the statistical model, and a brief treatment on the inclusion of additional explanatory and predictor variables. This later treatment segues into introducing the reader to the estimation of simultaneous within-group (Level 1) and between-group (Level 2) regressions and the use of group means (i.e., grand-mean centering) and within-group deviation scores (i.e., group-mean centering) for formal evaluation—with a detailed conceptual treatment of the necessity for and interpretation of associations and regressions among the same (nominal) variables at different levels of analysis. The remainder of Chapter 4 is devoted to alternative options when using multilevel models including parameter estimation (i.e., reduced and full information maximum likelihood; REML and ML) and empirical Bayes estimates for posterior means. Chapter 4 concludes with a brief example extending the previous examples in the chapter, introducing readers to the three-level model, detailing the multiple types of ICCs that can be calculated, as well as simultaneous regression across all three levels and the mathematics behind creating the correct grand- and group-mean centered variables for the predictor.

Chapter 5 extends the basic random intercept multilevel model to the development of the random slopes model. Continuing the example from Chapter 4, readers are introduced to the mathematics and equations behind the inclusion of a random slope as well as the covariance among random effects. The influence of centering predictor variables and the relationship (covariation) between random intercepts and slopes are treated, and readers are provided with information regarding the use of the random slope variance for examining the plausible range of values for random slopes. These set the backdrop for what many behavioral and social scientists are primarily interested in, the explanation of random intercepts and slopes. Here, the reader is introduced to the intercepts and slopes as outcomes models whereby heterogeneity in levels and slopes (i.e., random intercepts and slopes, respectively) can be explained and predicted. Specifically, readers are formally introduced to interactions among Level 1 predictors, Level 2 predictors, and cross-level interactions (i.e., moderation of a Level 1 predictor by a Level 2 predictor). The remainder of Chapter 5, similar to Chapter 4, focuses on issues pertaining to model estimation, including the advantages of REML for investigations of random effects, options, implications and interpretation of centering variables included as random effects, and finally a brief treatment on the parameterization of a random slopes model in the context of a three-level multilevel model.

Chapter 6 focuses on statistical testing and model specification. The brief treatment on null hypothesis significance testing provides the reader with a very clear understanding of calculating degrees of freedom and considerations for sample and unit size when testing effects in multilevel models. One important part of Chapter 6 is the more thorough and formal treatment of centering variables for within- and between-groups regression. Although the concept of simultaneous regression had been introduced, an example demonstrating the similarities and differences in estimates obtained from group-mean versus grand-mean centering are shown, as well as their mathematical equivalence. Deviance tests and methods and statistical tests for comparing the fit of nested models are presented, with an additional treatment specifically for deviance tests using the chi-bar distribution to obtain greater power when comparing models with respect to variance components. Finally, a considerable portion of Chapter 6 is devoted to model specification-the choice of a satisfactory model. The authors discuss the importance of both substantive and statistical considerations when specifying multilevel models and provide a rather comprehensive set of principles designed to help end users make practical, logical, and well-informed decisions regarding the inclusion (or exclusion) of model parameters. They supplement these principles with discussion of building the multilevel model. This discussion guides readers through a logical set of considerations for specifying the Level 1 equation for adequately modeling within-group variability, the inclusion of fixed followed by random effects in a sequential model-building process, followed by the development of the Level 2 equation to include predictors as moderators of Level 1 fixed effects and to account for variance via the reduction of random effects.

Chapter 7 focuses on the concept of explained variance in multilevel models. The problems of considering variance explained, similar to traditional multiple regression, is discussed with examples of how level-specific variance components (i.e., residual variance and random intercept) can exhibit peculiar properties (e.g., negative variance explained) depending on the specification of the parameters included and the (im-)balance of the data. Snijders and Bosker introduce the proportional reduction in prediction error (i.e., mean square error) as an alternative way to gauge variance explained. They provide examples of how to calculate this for variance at Level 1 and also discuss the additional considerations required for this calculation in a three-level framework. These concepts are further extended, albeit without empirical examples, to changes in variance components and random effects when predictors at different levels of analysis are included, and how predictors at one level of analysis can usurp variance at a different level of analysis.

Chapter 8 discusses heteroscedasticity, or models with heterogeneous variance components and random effects. The notion of variability in the residual variance is discussed in terms of possible differences in within-group variation that can reflect differences in systematic within-group variation and differences in measurement error. The advantages of heterogeneous variances models is further discussed in terms of their statistical importance, as failing to take into account systematic differences in variance components can result in incorrect standard errors for fixed effects, as well as their practical and theoretical importance in that group differences in variability could have important substantive meaning. Examples cover the inclusion of both categorical and continuous predictors, as well as within-group and between-group (Levels 1 and 2, respectively) predictors.

Chapter 9 details the estimation of multilevel models in the presence of missing data. The chapter starts with defining the classic types of missing data (missing completely at random [MCAR], missing at random [MAR], and missing not at random [MNAR]), and introducing the reader to some general implications of missing data and why it is important to consider analytically. In terms of handling missing data for analysis, two primary methods are treated: full information maximum likelihood (FIML) estimation and imputation. A very readable introduction to the concept of imputation is provided, with reference to more classic treatments of imputation and missing data given the considerable research into this topic over the past decade (e.g., Little & Rubin, 2002). A series of examples using simulated data serves to demonstrate the effect of missing data and listwise deletion versus single and multiple imputation approaches on parameter estimates in multilevel models, detailing the advantage of multiple imputation methods for reducing bias in estimates and standard errors. Finally, multiple imputation by chained equations, utilizing the available information and posterior distributions of the data, is detailed as an alternative imputation strategy, and considerations regarding the choice of imputation strategy are discussed.

Chapter 10 discusses the assumptions of the multilevel model, including the independence of random effects and Level 2 residual across groups, the independence of Level 1 and Level 2 residuals, and residuals having a mean of 0 and with a normal distribution and constant variance. This sets the stage for theoretical and practical treatment of best practices when employing multilevel models including the disambiguation of within- and between-groups regression via group- and grand-mean centering, the specification of the fixed and random portions of the multilevel model, and thorough treatments with empirical examples for inspecting and checking assumptions regarding the distribution of Level 1 and Level 2 residuals. With respect to the inspection of residuals, the reader is shown how to use normal probability plots and examine sources of heteroscedasticity in Level 1 residuals as a function of theoretically important predictors, as well as the use of influence diagnostics to examine how Level 2 units might differentially impact parameter estimation.

Chapter 11 provides the first introduction to issues regarding designing multilevel studies. Following a brief treatment of statistical power for single-level designs, issues specific to multilevel studies are discussed, such as the difference between using a two-stage (multilevel sample) versus a simple random sample with the same number of observations, cost-benefit considerations of the number of Level 1 and Level 2 units sampled both financially and statistically, considerations for Level 1 and Level 2 sample sizes for multilevel regression, and the role of the intraclass correlation in parameter estimation and power calculations. The majority of the chapter focuses on the role of Level 1 and Level 2 cluster sample sizes, and the ICC for power of examining mean differences in fixed effects and cross-level interactions. Brief treatment is given to power considerations for random effects and variance components given the less-developed empirical basis for such parameters.

Chapter 12 provides brief overviews of alternative methods and models that can be incorporated within the multilevel framework. Here, readers are introduced to Bayesian estimation, which had been touched on previously in the book. Alternative standard error estimation, specifically the sandwich standard errors or cluster robust standard errors, which provide adjustments in the event of heteroscedastistic residuals are described in detail as well. Finally, readers are introduced to latent class models as an extension of the multilevel model considering that Level 2 groups might further represent distinct categories or *latent classes.* Practical issues, such as assessing model fit using indexes that penalize for additional parameters (e.g., Bayesian Information Criterion), as opposed to deviance testing are introduced given the potential for considerable increase in parameter estimation.

Chapter 13 introduces the reader to the concept of imperfect hierarchies, or cross- or multiple classification structures. Three types of models are explicitly examined: cross-classification models, multiple membership models, and multiple membership multiple classification models. The estimation of cross-classification models introduces the use of a crossed random effect to account for additional dependency introduced by cross-classification. The estimation of multiple membership models introduces the notion of proportionally weighting of the variance components with respect to the time a Level 1 unit was nested within each of the discrete Level 2 units. Finally, the multiple membership multiple classification model is introduced as a simple extension of the previous two models to incorporate both crossed random factors and proportional weighting of random effects for dealing with changes in nesting over time.

Chapter 14 is a new addition to the book and introduces the concept of survey weights to increase precision of model estimates and reduce bias regarding inferences about a population. The reader is introduced to model-based and designed-based approaches to statistical inference. The former attempts to incorporate elements of study design into the model, whereas the latter incorporates weights based on sampling probabilities and are quite common in the world of survey methodology. The two primary weights introduced are sampling or survey weights, which take into account the probability of a unit being included in the sample, and precision weights, which weight observations so that observations with higher residual standard deviations have less influence (lower weight). The remainder of Chapter 14 is devoted to treatment of practical considerations when using model-based versus design-based approaches to analysis, and approaches considering the inclusion of weights and differences that emerge when using model- versus design-based analyses. Finally, the chapter concludes with a series of empirical examples demonstrating differences (and ways to compare such differences; e.g., Asparouhov, 2006) in parameter estimation based on design-based (weighted) and model-based (unweighted) analyses of data, and general considerations of creating weights with multilevel data.

Chapter 15 provides a thorough treatment of the multilevel model for longitudinal data. The application of the multilevel model to repeated measures data with both fixed- and variable-interval occasions is presented, demonstrating the virtues of multilevel analysis for treating time more flexibly. Also covered are the estimation of random slopes and the inclusion of predictors for examining differential change (i.e., individual differences in change) and variance accounted for in random effects. The notion of more complex, nonlinear change functions is examined through the use of higher order polynomial functions, piecewise linear models for discontinuous change, and fixed-knot spline models. Finally, a brief treatment of the inclusion of covariates, both time-invariant and time-varying, in longitudinal multilevel models is provided. One thing to note is that, compared with previous chapters, less detail is given to the development of fixed and random portions of the longitudinal models and considerations for the optimal parameterization of each.

Chapter 16 provides a brief treatment of multivariate multilevel models. Here, the reader is introduced to the multivariate model as an extension of the univariate model, allowing for greater flexibility in terms of the simultaneous prediction of multiple outcome variables, direct comparison of the magnitude of the effects of predictors on outcomes, examination and explanation of correlations among outcomes at different levels of analysis (i.e., within- and between-group). Although no empirical examples are provided, development of both random intercept and random slope models within the multivariate multilevel framework is provided, which gives a sufficient introduction to the multitude of opportunities for new and complex questions to be answered through the flexibility of the multivariate framework.

Chapter 17 introduces, for the first time, the application of nonlinear multilevel models for use with discrete outcomes. The application of the hierarchical generalized linear model is presented for dichotomous outcomes (multilevel logistic regression), ordered categorical outcomes (multilevel ordered logistic regression), and discrete count data (multilevel Poisson and negative binomial regression). The majority of the chapter focuses on the multilevel logistic model and the special complexities involved compared with estimation of the general linear mixed model for continuous outcomes. A very accessible tutorial on the basics of logistic multilevel modeling from the ground up is provided with a focus on the logit function, the interpretation of parameter estimates, complexities involving the estimation of variance components and random effects (e.g., residual variance pertains to the outcome explicitly, not the probability of its occurrence), and different methods of estimation such as Laplace approximation and adaptive numeric quadrature for approximating maximum likelihood estimates. Further treatment is given to variance decomposition and variance explained in logistic multilevel models with focus on both the ICC and residual ICC, providing symmetry for readers interested in the application of logistic multilevel models and being able to reconcile similarities and differences with the estimation of continuous outcome models.

Finally, Chapter 18 provides readers with an overview of both more mainstream and more specialized software packages that support multilevel analysis. The dedicated multilevel analysis packages include HLM (Bryk, Raudenbush, & Congdon, 1986), MLwiN (Rabash & Woodhouse, 1995), as well as MIXOR and SuperMix (Hedeker & Gibbons, 1996a, 1996b). Additionally, many general-purpose statistics packages include options for multilevel analysis, including SAS (the VARCOMP, MIXED, GLIMMIX, and NLMIXED SAS PROCs), R (the 'nlme' [Pinheiro & Bates, 1995]) and 'lme4' [Bates, 2010] packages), Stata (the 'loneway,' 'xt,' 'xtreg,' 'xtpois,' 'xtprobit' modules), and SPSS (the VARCOMP and MIXED commands). Finally, multilevel analysis software is available for special purposes such as power analysis including PinT (Snijders & Bosker (1993), Optimal Design (Spybrook, Raudenbush, Congdon, & Martinez, 2009), and MLPowSim (Browne, Lahi, & Parker, 2009), multilevel analysis with latent variables including M*plus* (Muthén & Muthén, 2010) and Latent Gold (Vermunt & Magidson, 2005a, 2005b); and specialized packages for more complex applications (e.g., Bayesian estimation) including REALCOM and WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000).

EVALUATION AND CONCLUSION

The primary strengths of Snijders and Bosker's book is its comprehensive, very readable treatments of the concepts discussed, striking a good balance between text and equations, and its focus on thoughtful and sound application to real-world multilevel analytic issues. Readers and end users familiar with Snijders and Bosker's first edition of this book will find this updated version an improved resource given the expanded attention to consider new topics relevant to multilevel study designs and analysis (e.g., treatment of missing data, advanced topics in model estimation). New readers will find the text a very palatable yet comprehensive resource on multilevel modeling from theory to study design, to application of the modeling framework, and even the statistical packages available. The Web site accompanying the book (http://www.stats.ox.ac.uk/∼snijders/) is another good resource for readers, containing data and scripts for the examples covered in the book. It should be noted that the Web site and online materials are a work in progress, as scripts are currently available for MLwiN, R, and Stata only, although complementary supporting materials for HLM, SAS, and M*plus* are on the horizon.

New features of this second edition that are particularly noteworthy are the addition of the *overviews* and *glommaries* bookending each chapter. The overviews make it even easier for researchers to identify chapters, at a glance, that will provide them with the information relevant for their particular needs, and the glommaries, which summarize and define the relevant terms for each chapter, lend this book to being a fantastic quick reference for all issues related to multilevel analysis. The addition of the glommaries improves the stock of this book as a primary text for classes on multilevel analysis, as well as a general one-stop-shop desk reference for multilevel analysis.

Although this book does have a considerable number of strengths, there are also issues regarding certain aspects of the book readers should keep in mind. I do not consider these issues to be weaknesses, as much as limitations that come with realities of writing books. Whereas the chapters covering the basic aspects of multilevel analysis are quite detailed and comprehensive, some of the more advanced topic chapters should be viewed as more introductory or primers on these topics with examples of their application in the multilevel analysis framework. The authors, however, do reference seminal and current state-of-the-art literature on these more specialized topics such as missing data (e.g., Schafer, 1997; Schafer & Graham, 2002), survey weights (e.g., Pfefferman, 1993; Pfefferman, Skinner, Holmes, Goldstein, & Rasbash, 1998; Raghunathan, Lepkowski, van Hoewyk, & Solenberger, 2001), and Bayesian estimation (e.g., Jackman, 2009), which helps in identifying appropriate readings for more in-depth reference as each have their own well-developed empirical literatures.

Also, although the book does a good job of developing the symmetry between the application of the multilevel model to nested designs (i.e., students nested in classrooms) and longitudinal and repeated measures designs (i.e., observations nested within an individual), the empirical examples in Chapter 15 demonstrating the multilevel analysis of longitudinal data are somewhat more difficult to reconcile. The description of the design for the studies used in the empirical examples is somewhat unclear with respect to the exact number of observations available for analysis, and heterogeneity in age and birth cohort at baseline of the subsamples employed for analysis. These pieces of information are important for understanding how time or the basis for change for the fixed and random effects of the longitudinal model are developed and estimated. As such, the examples for their fixed occasion and variable occasion designs appear to be estimated capitalizing on the heterogeneity in age or cohort at the initial occasion and subsequent longitudinal follow-up on each participant to use a multiple cohort linkage approach (e.g., Miyazaki & Raudenbush, 2000) or what has also been termed *convergence* (Raudenbush & Bryk, 2002). The first empirical example, which employs a fixed occasions design, considers change in life satisfaction from age 55 to 60 and considers birth year as a Level 2 predictor, which takes into account cross-sectional differences as a function of age or cohort and allows for the disambiguation of age-related differences from aging-related change in life satisfaction, and is consistent with the authors', as well as others' (e.g., Curran & Bauer, 2011; Hoffman & Stawski, 2009) repeated assertion for disambiguating variability in predictors so between- and within-group or person regressions are distinguished. Subsequent models such as the examples considering variable occasion designs and more complex growth and change functions (e.g., polynomial, piecewise linear, and spline models) use age as the time basis for the models, but appear to use convergence models for examining change. Such models capitalize on the variability across individuals (in this case, with respect to age or cohort; e.g., Hofer & Sliwinski, 2001), and allow for more complex functions to be estimated that couldn't be estimated or observed for any given individual given a potentially limited number of observations. This has important implications for the interpretation of both fixed and random effects and what they really reflect (i.e., age differences vs. developmental change), which individuals are contributing to the estimating of model parameters, and considering model fit and variance explanation. The approach taken in the examples is not incorrect, per se, but is one of many approaches taken to analyze longitudinal data. It is, however, an issue that has received considerable attention and is still debated in the area of longitudinal analysis (Mehta & West, 2000; Miyazaki & Raudenbush, 2000; Molenaar, 2004), with recent calls for researchers to, at the very least, formally examine the convergence assumption empirically (Sliwinski, Hoffman, & Hofer, 2010). There is certainly a limit to the amount of material that can be treated in any given chapter, and dedicated texts on intensive repeated measures and longitudinal analyses are available (e.g., Hedeker & Gibbons, 2006; Singer & Willett, 2003; Walls & Schafer, 2006).

In conclusion, taken together, this book is a fantastic resource for self-instruction by end-user researchers, instructors of statistics and methods courses who are looking for a comprehensive and extremely comprehensible resource on multilevel modeling, and more advanced researchers and analysts who desire a resource to expand their multilevel analytic toolkit. Although this text focuses on empirical examples using nested data and longitudinal data being a special case of multilevel design (observations nested within individuals), it does do a good job of conveying the importance of attending to multilevel structures for both the nested data and repeated measures and being able to mentally toggle between the two. I find this to be an important feature as students new to thinking about multilevel data structures can sometimes struggle with understanding how the same model applies to substantively different designs. Furthermore, although there are equations throughout the chapters, the book is by no means "math-centric," which will help the book to seem less intimidating and appeal to a breadth of potential readers. Thus, the book should have broad appeal and utility to researchers and students in both the social and behavioral sciences.

REFERENCES

Asparouhov, T. (2006). General multilevel modeling with sampling weights. *Communications in Statistics: Theory and Methods, 35,* 439–460.

Bates, D. M. (2010). lme4: Mixed-effects modeling with R. Retrieved from http://lme4.r-forge.r-project.org/book/

Borsboom, D., Mellenbergh, G. J., &Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110,* 203–219.

Browne, W. J., Lahi, M. G., & Parker, R. M. A. (2009). *A guide to sample size calculations for random effects models via simulation and the MLPowSim software package.* Bristol, UK: University of Bristol.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). *HLM: Hierarchical linear modeling with the HLM/2L and HLM/3L programs.* Chicago, IL: Scientific Software International.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62,* 583–619.

Hedeker, D., & Gibbons, R. D. (1996a). MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine, 49,* 157–176.

Hedeker, D., & Gibbons, R. D. (1996b). MIXREG: A computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine, 49,* 229–252.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis.* New York, NY: Wiley.

Hofer, S., & Sliwinski, M. (2001). Understanding ageing: An evaluation of research designs for assessing the interdependence of age-related changes. *Gerontology. 47,* 341–352.

Hoffman, L., & Stawski, R. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development, 6,* 97–120.

Jackman, S. (2009). Bayesian analysis of choice making. In *Bayesian analysis for the social sciences.* Chichester, UK: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5,* 23–43.

Miyazaki, Y., & Raudenbush, S. W. (2000). A test for linkage of multiple cohorts from an accelerated longitudinal design. *Psychological Methods, 5,* 44–63.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2,* 201–218.

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

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data.*International Statistical Review, 61,* 317–337.

Pfefferman, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of Royal Statistical Society, Series B, 60,* 23–40.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in nonlinear mixed-effects models. *Journal of Computational and Graphical Statistics, 4,* 12–35.

Rabash, J., & Woodhouse, G. (1995). *MLn: Command references.* London, UK: Multilevel Models Project, Institute of Education, University of London.

Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a series of regression models. *Survey Methodology, 27,* 85–96.

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

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

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London, UK: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Sliwinski, M., Hoffman, L., & Hofer, S. (2010). Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. *Research on Human Development, 7,* 45–60.

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level models. *Sociological Methods and Research, 22,* 342–363.

Spybrook, J., Raudenbush, S. W., Congdon, R., & Martinez, A. (2009). *Optimal design for longitudinal multilevel research: Documentation for the Optimal Design software, version 2.0.* Ann Arbor, MI: University of Michigan.

Vermunt, J. K., & Magidson, J. (2005a). *Latent GOLD 4.0 user's guide.* Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2005b). *Technical guide for Latent GOLD 4.0: Basic and advanced.* Belmont, MA: Statistical Innovations.

Walls, T. A., & Schafer, J. L. (2006). *Models for intensive longitudinal data.* New York, NY: Oxford University Press.