

# Advances in Multilevel Modeling for Educational Research: Addressing Practical Issues Found in Real-World Applications

[tcrecord.org/PrintContent.asp](http://tcrecord.org/PrintContent.asp)

reviewed by [Manuel S. González Canché](#) — February 21, 2017

**Title:** Advances in Multilevel Modeling for Educational Research: Addressing Practical Issues Found in Real-World Applications

**Author(s):** Jeffrey R. Harring, Laura M. Stapleton, & S. Natasha Beretvas (Eds.)

**Publisher:** Information Age Publishing, Charlotte

**ISBN:** 1681233282, **Pages:** 414, **Year:** 2015

[Search for book at Amazon.com](#)



It is often the case that research questions in education involve units of analysis that can be naturally grouped or placed within hierarchical or multilevel configurations. This type of grouping is referred to as nesting. It results in the exposure of the lowest-level units of analysis to common environments that are likely to impact their behaviors, outcomes, or levels of performance. These lowest-level units are commonly referred to as level-1 units. Typical examples of these units are students *nested* within classrooms. Classrooms are the nesting structure, constituting a second-level unit or level-2 unit. Following this rationale, researchers could further model level-3 units consisting of classrooms (comprised of students) nested within schools. Notably, this logic can be further expanded to higher order levels. The successful identification of units situated at different levels prompted the development of techniques designed to model this phenomenon. These techniques are known as *multilevel modeling*.

Jeffrey R. Harring, Laura M. Stapleton, and S. Natasha Beretvas edited the volume *Advances in Multilevel Modeling for Educational Research: Addressing Practical Issues Found in Real-World Applications*. It offers state of the art procedures designed for addressing issues faced in daily research requiring multilevel modeling. For example, classical multilevel modeling was designed with simpler designs in mind where students are assigned to classrooms with a single teacher. Whenever there are multiple teachers assigned to a single classroom, the design becomes fuzzier in that any variation in student outcomes is affected by the potential contributions of several teachers rather than just one. Similarly, it may be that a single teacher influences students in different classrooms. This type of problem is not just a reflection of *real-world* situations, but is also the result of greater data availability that enables the potential for modeling these same real-world problems. For example, assume we have access to a sharp design with only one teacher per classroom. Even in this simple case, some students may still have access to private tutors or other extracurricular activities that directly or indirectly affect their academic performance (e.g., music, chess, etc.), therefore introducing another unaccounted source for variation. It may also be similar in the case of a single-teacher, single-classroom design where students neither participate in extracurricular activities, nor have access to private tutors. In this case, physical classroom configuration or the organic formation of friendships among students from different classrooms could impact academic performance. In this latter case, friendship effects could potentially be modeled relying on social network analysis principles, as illustrated in Chapter Twelve by Sweet and Zheng.

Given that this book builds upon traditional multilevel modeling principles and exposes readers to real-world issues faced in multilevel modeling, its content demands a strong basic understanding of the notions behind multilevel modeling. The volume's level of sophistication means that its content may not be suitable for readers expecting a gentle introduction to multilevel or hierarchical level modeling. Nonetheless, the authors provide readers with clear

explanations of the rationale behind their models and show examples that guide them through outcome interpretation and relevance. What follows in this review is not intended to serve as a summary of the text. Instead, it discusses the book sections, highlights their goals, and analyzes the methods mentioned in each chapter.

The book is organized into three sections: “Methodological Issues in Multilevel Models,” “Handling Longitudinal Data Complexities: Cross-classified and Random Effects Latent Variable Models,” and “Causal Inference and Advance Psychometric Models.” The first section contains four chapters. In Chapter One, Bauer and Curran call attention to the ways error measurements may cause biased inferences. The authors highlight common practices employed in longitudinal panel studies. These consist of (a) adding together ordinal items without implementing outcome calibration procedures, or (b) calibrating outcome measures in longitudinal studies with information obtained at a single point in time. In the latter approach, the possibility remains that findings are still biased even with the sophisticated calibration of outcome measures in cases when such calibration measures only consider baseline information. The authors discuss pros and cons of three calibration procedures and finish by indicating that more research is needed. They also believe that additional emphasis should be placed on strengthening the alignment between outcome measurement and modeling in longitudinal research.

Chapter Two, by Hox, van Buuren, and Jolani, deals with missing values at the group level. This is a problem that places the entire multilevel modeling procedure at risk of serious bias. For example, the authors present the scenario where a single missing variable at the teacher level leads to the exclusion of the entire group of students when deleting cases listwise. Rather than deleting observations with missing information or applying group mean imputation techniques, Hox, van Buuren, and Jolani discuss full information maximum likelihood (FIML) and multiple imputation (MI) in the context of multilevel data. The authors also highlight challenges and best practices such as the inclusion of auxiliary variables during data collection that may serve to predict who is likely to drop from the study, becoming part of the *missing data group*. These auxiliary variables consist of asking participants about the perceived likelihood of their continuation in the next study’s follow-up. This is a practice that is further highlighted in other chapters throughout the book.

In Chapter Three, Stapleton, Haring, and Lee discuss sampling weight issues that are particularly relevant for researchers using surveys intended to be nationally representative. This includes surveys like those collected by the National Center for Education Statistics. Similar to the discussion presented in Chapter Two, Stapleton, Haring, and Lee are also interested in sampling weight adjustment for intermittent nonresponse and attrition. Specifically, after presenting a summary of panel weighting approaches typically employed for nonresponse, they focus on alternative methods such as FIML and MI with auxiliary variables. Notably, the authors close this chapter by indicating that the current state of conceptual and methodological knowledge is not currently at par with the inherent challenges presented by missing data and attrition in panel surveys. The first section concludes with Chapter Four where O’Connell, Yeomans-Maldonado, and McCoach discuss residual diagnostics and model assessment. Related to this point, it is worth noting that I previously tested for spatial dependence of residuals obtained from multilevel modeling. In this case, I found that multilevel modeling was unable to correct for this problem when nesting units of analysis at the state level (González Canché, 2014). This finding is important considering that the main concern raised by O’Connell, Yeomans-Maldonado, and McCoach is that residual analyses remain practically unused in the multilevel framework. This may negatively influence the sorts of inferences reached. Accordingly, the authors’ main contribution is offering a three-staged approach where the identification of the best level-1 model is prioritized before examining [m1] coefficient weights.

The second section begins with Chapter Five where Jiao, Kamata, and Xie deal with issues of cross-classification. This pertains to situations where the number of dependencies is the result of exposure to more than one condition. For example, students could be clustered with several teachers or clustered both in classrooms and extracurricular activities. The authors introduce a multilevel cross-classified testlet model that accounts for cross-classification issues. The results from their proposed model show better fit than seven other models that do not deal with this cross-classification problem. In Chapter Six, Asparouhov and Muthén use Bayesian methods to go beyond maximum likelihood (ML) and weighted least squares (WLS) models that are incapable of fitting cross-classified structural models and random loading models. Perhaps the most important contribution of this Bayesian approach is that the Bayesian methodology is robust against non-positive definite matrices, singular-variance covariance matrices, or negative residual variances obtained when including a large number of random effects. This is different from ML and WLS.

Chapter Seven, by Mehta and Petscher, introduces an extension of structural equation modeling (SEM) that is capable of fitting SEMs at  $n$  number of levels. In this case,  $n$  stands for an arbitrary number of nesting structures. The authors are particularly careful to consider the multiple *moving parts* that may impact inferences. These parts include teachers in charge of multiple classrooms, students moving across classrooms over time, and a combination of both of these factors. In this view, the methodological approach is similar to the procedure used by González Canché and Rios-Aguilar (2015). In this instance, we were measuring the effects of classmates over a two-year period in a community college setting. In this study, we employed network analysis principles to capture the average effect of all classmates over time. In Chapter Seven, Mehta and Petscher take a similar approach, but model it under the SEM approach. This choice enables the incorporation of 11 levels. Another aspect worth noting is that the authors rely on free software (e.g., freeware) for their models and use a set of algorithms contained in the xxM package designed for the R language.

Harring, Beretvas, and Israni (Chapter Eight) present an approach to modeling the multilevel cross-classification aspects of longitudinal data requiring measurement of changes in students' attributes, skills, or performance. They also gather information on factors affecting these students' prospects of success. These factors include teachers, schools, and neighborhoods. This chapter is another example where its great amount of data illuminates the necessity for analytic techniques that purposefully attempt to model the myriad number of real-world factors affecting student performance. In Chapter Nine, Beretvas, Murphy, and Gaertner use multilevel measurement models to estimate rater effects after accounting for item difficulty, person ability, and rater severity. Similar to Chapter Eight, Beretvas, Murphy, and Gaertner depict the need to model data structures that reveal real-world issues. Specifically, the authors discuss similarities and differences among three versions of multilevel measurement models that can handle rater effects to show impact assessment, differential item functioning, and differential rater functioning.

The final section of the book includes three chapters. In Chapter Ten, Yang & Seltzer propose an approach to handling measurement error in multilevel modeling. In this approach, observed categorical variables are assumed to capture latent predictors. The authors use a two-stage approach where the first stage consists of specifying and imputing sets of plausible values of latent measures that are treated as latent variable predictors. Subsequently, they use these newly estimated variables within a multilevel model at a later stage. Unsurprisingly, the authors note that the success of this approach depends on the specification of a model that feasibly captures the assumed latent structure of the predictors, which requires further simulation and research.

In Chapter Eleven, Kim, Steiner, and Lim rely on propensity score models to discuss how the lack of overlap in

individuals' propensities toward treatment within clusters can be addressed. They show that before creating comparable cases by moving participants across clusters to deal with the lack of overlap in probabilities of receiving treatment, the authors need to evaluate heterogeneity in the selection process. Kim, Steiner, and Lim convincingly show mechanisms implemented to remove heterogeneous selection bias and obtain less biased estimates of treatment effects. Specifically, they show a propensity score modeling strategy that identifies clusters with similar propensities regarding the selection process. It then matches treated and control units only within the homogenous classes of clusters.

In the twelfth and final chapter, Sweet and Zheng rely on social network analysis principles to incorporate individual- (or node-) level, network-level and dyad- (or relation-) level covariates. Notably, the models presented by Sweet and Zheng deal with relationship (or tie) formation. Researchers interested in exploring how ties or relationships affect students' outcomes may benefit from a consideration of the network analysis procedures discussed in González Canché and Rios-Aguilar (2015).

Generally speaking, *Advances in Multilevel Modeling for Educational Research* is an excellent resource for applied researchers looking for state of the art procedures to model complex data structures using multilevel analysis. The authors' rich sets of equations constitute an important guide for applied researchers interested in building upon these approaches. The volume is also an excellent resource for statisticians, methodologists, and computer scientists interested in funding opportunities that pursue methodological issues in education research. The authors excel in depicting the next steps required to continue advancing the multilevel modeling world and therefore the potential for grant funding. Perhaps the most notable issue omitted from this book relates to the influence of physical space and distance in modeling dependence (see González Canché, 2014, 2016). This is the analysis of units that are clustered as a function of physical distance. Given the availability of geocoded data and the proposition that near things are more related than distant things (Tobler, 1970), consideration of space in the hierarchical modeling framework would enable integration of multilevel, network, and spatial analysis methodologies that may result in less biased inferences in education research.

## References

González Canché, M. S. (2014). Localized competition in the non-resident student market. *Economics of Education Review*, 43, 21–35.

González Canché, M. S. (2016). Assessing the illusion of independence: How spatial econometrics and network analysis can strengthen. *IHE Report*. Retrieved from <http://ihe.uga.edu/about/ihe-report-article/assessing-the-illusion-of-independence-how-spatial-econometrics-and-network/>

González Canché, M. S., & Rios-Aguilar, C. (2015). Critical social network analysis in community colleges: Peer effects and credit attainment. *New Directions for Institutional Research*, 2014(163), 75–91.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*,

46(sup1), 234–240.