Analyzing the Effects of the Absence of Levels in Multilevel Modeling using TIMSS Data

Breanna Morrison, Xinya Liang, Wen-Juo Lo – University of Arkansas

Abstract

Multilevel models are popular for handling data nested within multiple higher-level clusters. It provides substantial power and accuracy if the levels of data are properly specified. However, leaving out higher levels of nested data may result in biased standard errors and inaccurate variance estimates for lower levels[2,3].

The current study investigated how reduced models (i.e., ignoring a level of the nested structure) affects various outcomes in estimating students' math achievement from a variety of potential predictors. Using TIMSS 2011 and complementing with literature, an exploratory analysis was conducted to identify predictors to be used in the three-level *full* model. Then two reduced 2-level models, one with students nested within schools, and the other with students nested within teachers, were compared to the full model.

As expected, the fits of two reduced models were significantly worse than the full model. Further, estimates and standard errors tended to be overestimated at the second level while compared to the full model.

Introduction

In educational research, data are often hierarchically structured such as students are nested within classrooms. With this type of data, using statistical analyses that assume independence of the data points can cause inaccurate results [1,4,5]. Hierarchical linear modeling (HLM), as an alternative approach, allows for modeling data with multiple levels. For instance, a 3-level HLM model may have data consisting of students (level 1) nested within teachers (level 2) and teachers nested within schools (level 3). Using empirical data, the purpose of this study is to evaluate the impact of leaving any higher level out of the full model and how that affects the accuracy of model estimation.

Many studies have been looking at the impact of leaving out a level of data in a multilevel model[2,3,6]. These studies have typically found that leaving out a level in a model makes the model redistribute that variance into the other levels. Opdenakker & Van Damme (2000) used a 4-level educational model (i.e., student, class, teacher, and school) and found that when the highest level was ignored, the next highest level had inflated estimates of the variance. If a middle level was ignored, the level above the middle level had an underestimation of the variance it was accounted for while the level below the ignored level gave an overestimation of the variance it was accounting for. The regression coefficients of the levels affected by the missing level also tended to be inaccurate. This way of the variance being redistributed followed the patterns of other studies mentioned[2,6].

Research Question: How will leaving out a level from what should be a three-level model affect the estimates and variances calculated?

Method

This research used the Trends in International Mathematics and Science Study (TIMSS) data in 2011. TIMSS collects data at the classroom and school level making it possible to run analysis on a 3-level model.

Using 4th grade American students with respective teachers and schools, three models were evaluated:

1) Three-level full model with levels of students, classes, and school categories
2) Two-level reduced model with students within classes
3) Two-level reduced model with students within school categories.

First, students who had multiple teachers listed were not included. Second, to obtain sufficient sample sizes at the class level, schools were combined to make 8 school categories based on percent of economically disadvantaged students (0-10%, 11 to 25%, etc.) and area type (urban, rural, etc.). Overall, there were 2,676 students, 189 classes, and 8 school categories.

After the exploratory analysis, we identified three variables at level 1, and one variable for level 2 and level 3, respectively. The reduced models used the same predictors.

Level-1 Predictors

- [Books] Amount of books in home
- [Computer] How often student uses a computer not at school or home
- [Confidence] Confidence in Mathematics

Level-2 Predictor

- [Prerequisites] How much does student's prerequisites limit teaching ability

Level-3 Predictor

- [Curriculum] How well do teachers teach to the curriculum

Level 1 Model: $\text{MathAchievement}_{ijk} = \pi_{0jk} + \pi_{1jk}*(\text{Books}_{ijk}) + \pi_{2jk}*(\text{Computer}_{ijk}) + \pi_{3jk}*(\text{Confidence}_{ijk}) + e_{ijk}$

Level 2 Model: $\pi_{0jk} = \beta_{00k} + \beta_{01k}*(\text{Prerequisites}_{jk}) + r_{0jk}$
$\pi_{1jk} = \beta_{10k} \quad \pi_{2jk} = \beta_{20k} \quad \pi_{3jk} = \beta_{30k}$

Level 3 Model: $\beta_{00k} = \gamma_{000} + \gamma_{001}(\text{Curriculum}_{k}) + u_{00k}$
$\beta_{01k} = \gamma_{010} \quad \beta_{10k} = \gamma_{100} \quad \beta_{20k} = \gamma_{200} \quad \beta_{30k} = \gamma_{300}$

Result

	Model 1 (Full model)		Model 2 (Students within Classes)		Model 3 (Students within Schools)	
Variables	**Coefficient**	***SE***	**Coefficient**	***SE***	**Coefficient**	***SE***
Level 1						
Intercept (γ000)	552.15**	4.40	551.49**	2.56	552.54**	5.00
Books (γ100)	9.21**	.98	9.21**	.98	12.65**	.98
Computer (γ200)	9.27**	.97	9.27**	.97	10.99**	1.01
Confidence (γ300)	11.91**	.52	11.91**	.52	11.33**	.53
Level 2						
Prerequisite (γ010)	-12.16*	3.64	-17.81**	4.36		
Level 3						
Curriculum (γ001)	-28.94*	7.48			-31.32*	8.35

***Note*: The estimates are for the fixed effects for the intercepts.**
*** p < .01, ** p < .001**

The fits of both reduced models were significantly worse than the full model (Model 2 vs. 1, χ^2_{diff} (7) = 46.46, p < .001, and Model 3 vs. 1, χ^2_{diff} (7) = 138.45, p < .001).

- ❑ Model 2 vs. Model 1 (Coefficients and standard errors)
  - Level-1: Estimates were extremely similar.
  - Level-2: Estimates were different.
- ❑ Model 3 vs. Model 1 (Coefficients and standard errors)
  - Level-1: "Confidence in mathematics" remained similar, but both "Amount of books in home" and "Computer use not at home/school" were different between models.
  - Level-3: Estimates were different.

Model	Deviance (# of parameters)	Variance Component Level 1	Level 2	Level 3
1	29054.62 (9)	2734.86 (78.48%)	631.70 (18.13%)	118.02 (3.39%)
2	29101.08 (2)	2738.36 (72.62%)	1032.57 (27.38%)	
3	29193.07 (2)	3199.08 (94.40%)		189.76 (5.60%)

***Note*: Percent of total variability at the level is listed in the parenthesis and was calculated from the variance components.**

In terms of variance components, the accuracy of estimation depends on the number of higher level units. Not surprisingly, the estimates vary across three models.

Conclusion

Our findings consistent with previous studies which leaving out levels makes not only fitting worse but also differing estimates and standard errors. Simulation studies might be a good way to investigate this in a more controlled situation.

Limitations:

- ❑ There were only 8 school categories at level-3 and no significant variation among the slopes at level 3. Due to no significant slopes in this study, we could not evaluate how missing levels may affect the slopes being estimated.
- ❑ The predictors were on the single likert-type scale rather than psychometric scales which could make the comparisons between the models work in more erratic ways.

Reference

1. Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, 41(2), 372-384.
2. Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*(1), 129-149.
3. Opdenakker, M. C., & Van Damme, J. (2000). The importance of identifying levels in multilevel analysis: an illustration of the effects of ignoring the top or intermediate levels in school effectiveness research. *School Effectiveness and School Improvement, 11*(1), 103-130.
4. Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research & Evaluation, 7*(1), 1-3.
5. Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd Edn*. Thousand Oaks, CA: Sage Publications, Inc.
6. Tranmer, M., & Steel, D. G. (2001). Ignoring a level in a multilevel model: evidence from UK census data. *Environment and Planning A, 33*(5), 941-948.

UNIVERSITY OF ARKANSAS
College of Education & Health Professions
Educational Statistics & Research Methods

2017 Modern Modeling Methods Conference, Storrs, CT