Controlling Multi-level Error in Single-level Studies: A Design-based Approach

Authors: Wang Yang, Wen Zhonglin, Fu Yuanshu, Zhonglin Wen

Date: 2022-03-01T00:00:00+00:00

Abstract

Owing to sampling design considerations, multilevel data structures are not only present in multilevel research but also prevalent in single-level research, necessitating the control of multilevel errors within single-level analyses. In such contexts, multilevel linear models fail to confer advantages and instead introduce complications stemming from model complexity. Design-based methods are comparatively simpler, more efficient, and robust, rendering them more suitable for single-level research scenarios encompassing multilevel errors. Following a detailed exposition of design-based methods and their advantages, this study employs empirical data examples to demonstrate the efficacy of design-based methods in controlling multilevel errors in single-level research and provides methodological selection recommendations for applied researchers.

Full Text

Controlling Multilevel Errors in Single-Level Studies: Design-Based Methods

Wang Yang[1], Wen Zhonglin[2], Fu Yuanshu[3]

[1] School of Public Administration, Guangdong University of Finance, Guangzhou 510521, China
[2] Center for Studies of Psychological Application / School of Psychology, South China Normal University, Guangzhou 510631, China
[3] School of Education, Zhaoqing University, Zhaoqing 526061, China

Abstract

Due to sampling design considerations, multilevel data structures exist not only in multilevel research but also widely in single-level studies, necessitating the control of multilevel errors within single-level analyses. In such contexts, hierarchical linear modeling (HLM) offers no particular advantage and instead introduces complications due to its model complexity. Design-based methods (DBM) are relatively simpler, more efficient, and more robust, making them better suited for single-level research contexts containing multilevel errors. After providing a detailed introduction to DBM and their advantages, this paper uses a data example to demonstrate the effectiveness of DBM in controlling multilevel errors in single-level studies and offers methodological recommendations for applied researchers.

Keywords: single-level research; multilevel data; hierarchical linear model; design-based methods

1. Introduction

In social science fields such as psychology, education, and management, using hierarchical linear models (HLM) to analyze multilevel data is common practice. However, few researchers realize that single-level studies also face problems with multilevel data. The most typical situation arises when, limited by human and financial resources, many studies do not employ random sampling but instead use cluster sampling or multi-stage sampling. For example, when several schools are selected and some or all students within them are sampled, the resulting data often have a multilevel structure (Huang, 2016), meaning that observations from the same school exhibit some degree of dependence. In such cases, although the researcher' s goal is only to analyze variable relationships at the student level without examining school-level variables (higher-level variables), the regression error term becomes contaminated with multilevel variation (i.e., multilevel errors). If the usual approach of ordinary least squares regression (OLS; i.e., familiar single-level regression analysis) is employed, it may cause bias in parameter estimation, particularly in the estimation of standard errors for regression coefficients (McNeish, 2014a). The literature refers to this unplanned multilevel structure as *incidental clustering* (McNeish & Wentzel, 2017), distinguishing it from deliberately designed multilevel structures.

What constitutes a better approach for handling such situations where single-level research encounters multilevel data? This question has received little attention. Although HLM can be used, it has numerous limitations. Design-based methods (DBM), by contrast, are tailor-made for such problems, allowing researchers to treat multilevel data as if conducting single-level analysis. This paper discusses the limitations of HLM, introduces the principles and advantages of DBM, demonstrates through an applied example how DBM controls multilevel errors in single-level studies, and finally provides methodological recommendations for applied researchers.

2. Limitations of HLM in Handling Multilevel Errors in Single-Level Studies

When single-level research results in data that actually contain a multilevel structure due to sampling design, HLM naturally comes to mind. However, using HLM in such situations has the following limitations.

2.1 Weakened Advantages

As is well known, HLM' s handling of multilevel data actually includes two functions: (1) controlling multilevel errors to ensure accurate standard error estimation for fixed effects of variables, and (2) analyzing random effects and making statistical inferences about group-specific effects. Of these two functions, the former is not exclusive to HLM, while the latter represents HLM' s unique advantage. However, in research focusing on single-level variable relationships, the nature of random effects is not of concern; the research task is simply to examine single-level variable relationships while controlling for multilevel errors —that is, to achieve function (1) only. Consequently, HLM' s advantages are not realized. Moreover, reviewing previous literature reveals that even in typical multilevel research, the purpose of using HLM is usually only to analyze fixed effects, with random effects treated as nuisance factors or secondary concerns (McNeish, 2014a).

2.2 Amplified Disadvantages

In single-level research, HLM' s advantages cannot be realized, while its disadvantages are further amplified. First, HLM has multiple assumptions about random effects. Regardless of whether researchers are interested in random effects, these assumptions must still be satisfied in HLM analysis. However, in the context of single-level research, because the design is not intentionally multilevel, the level-2 sample size is typically insufficient, making HLM assumptions harder to satisfy and verify. Table 1 summarizes the potential estimation biases that HLM may produce for level-1 effects (i.e., single-level variable relationships) when assumptions about random effects are violated.

Although these potential biases are limited to specific simulation contexts and do not indicate that assumption violations cause problems under all conditions, they at least demonstrate the risks associated with failing to satisfy random effects assumptions under certain conditions.

Second, to avoid estimation bias caused by omitting necessary random effects (McNeish, 2019), some researchers recommend including as many random effects as possible in HLM modeling (Barr et al., 2013). However, this approach also has problems: even if the model is theoretically identifiable, the information in the current data may not support an overly complex model (i.e., overparameterization), leading to convergence problems such as slow convergence or failure to converge (Bates et al., 2015). These problems are even more likely when sample sizes at each level are small.

Third, for a single-level study, using HLM is like using a sledgehammer to crack a nut—it unnecessarily complicates the problem. Theoretically, researchers must carefully consider whether and which random effects to include; in model specification, variables and model settings must be specified separately for between-group and within-group levels; in results presentation, because mathematical formulas have nested relationships (e.g., level-2 intercept and slope equations nested within level-1 equations) and require multiple subscripts to distinguish variables and error levels, HLM' s classical notation system is far more complex than single-level regression (as directly reflected in the popular HLM software); HLM result interpretation is also more complicated than single-level regression.

2.3 Summary

When single-level research results in data that actually contain a multilevel structure due to sampling design, using HLM for statistical analysis serves only to control multilevel errors. Its core advantages cannot be realized, yet researchers must endure the difficulties brought by HLM' s complexity. Therefore, HLM is not an ideal choice in such situations.

3. Design-Based Methods

Since the research question of interest remains a single-level problem, is it possible for single-level researchers to handle such actually multilevel-structured data more simply within their familiar single-level framework? DBM provides a solution to this problem. The reason why conventional OLS goes wrong in handling multilevel data is that the regression error term contains multilevel variation, leading to misestimated standard errors and consequently biased significance tests. HLM solves the multilevel error problem by decomposing variables and regression errors across different levels; DBM adopts a different strategy—correcting rather than decomposing the regression error term—which can also accurately estimate standard errors. Its essence is single-level regression that controls for multilevel errors. This special type of single-level regression fully accounts for the characteristics of complex sampling design data (i.e., multilevel structure) and is therefore called a design-based method.

DBM has gradually attracted attention from psychology researchers in the past five to six years. Numerous simulation and empirical studies have shown that they can effectively handle multilevel-structured data (e.g., Huang, 2016; McNeish & Stapleton, 2016) and have unique advantages over HLM, particularly in single-level research contexts containing multilevel errors. Through reviewing and synthesizing literature on multilevel analysis across various disciplines, three DBM approaches are particularly recommended.

3.1 Cluster-Robust Standard Errors

The first common DBM is the cluster-robust standard errors (CRSE) method, which corrects OLS standard errors by altering the calculation of regression coefficient variance (i.e., the square of standard errors). For non-nested data, standard single-level regression can be represented by the equation $Y = X\beta + \varepsilon$, where Y is an $n \times 1$ vector of the dependent variable, β is a $p \times 1$ vector of regression coefficients, X is an $n \times p$ design matrix, and ε is an $n \times 1$ vector of residuals. The variance of regression coefficients β can be estimated by:

$$\widehat{\text{var}}(\beta) = (X^T X)^{-1} X^T \widehat{\text{var}}(\varepsilon) X (X^T X)^{-1}$$

If the residual term ε is assumed to be independently and identically distributed (i.i.d., following a normal distribution with mean 0 and variance σ^2), then formula (1) can be simplified to:

$$\widehat{\text{var}}_{\text{OLS}}(\beta) = \sigma^2 (X^T X)^{-1}$$

The square root of the diagonal elements yields the coefficient standard errors based on OLS (McNeish et al., 2017). If the data have a multilevel structure, the residual term contains multilevel variation, resulting in heteroscedastic residual variance that violates the identical distribution assumption; within-group residual correlation is not zero, violating the independence assumption. CRSE corrects standard errors for both non-identical distribution and non-independence problems. For the non-identical distribution problem, the simplified formula (2) assumes homoscedasticity and misestimates regression coefficient variance, consequently misestimating standard errors. Therefore, CRSE abandons this simplified formula and adopts the complete formula for $\widehat{\text{var}}(\varepsilon)$, i.e., formula (1). For the non-independence problem, CRSE addresses it by replacing $\widehat{\text{var}}(\varepsilon)$ in formula (1) with $\sum_{j=1}^{J} X_j^T \hat{\varepsilon}_j \hat{\varepsilon}_j^T X_j$, which can be understood as aggregating individual-level residual terms within each group to form j new group-level residual terms for calculating standard errors. Data dependence occurs at the within-group level, while the between-group independence assumption is typically reasonable, so the above correction solves the independence problem. For more technical details about CRSE, see Cameron and Miller (2015).

The main prerequisite for using CRSE is that observations from different groups are uncorrelated (McNeish et al., 2017). This method' s characteristic is that it can be used not only to control multilevel errors in single-level research but also to directly analyze fixed effects of level-2 variables in dedicated multilevel studies. Moreover, it easily outputs R^2 statistics as effect size indicators, whereas calculating effect sizes in HLM is much more complex. Additionally, extensive statistical software supports this method (e.g., Mplus, R, SPSS, SAS, and Stata), providing conditions for broad application.

A major limitation of CRSE is that it only corrects standard errors of regression coefficients, not the coefficients themselves. The accuracy of regression coefficient estimation depends on the parameter estimation method originally used in single-level regression (e.g., OLS or maximum likelihood estimation).

This may pose problems in longitudinal data or data with extremely high intraclass correlation coefficients (ICC) (McNeish et al., 2017). Fortunately, at least in cross-sectional studies, even OLS provides accurate regression coefficient estimates (Huang, 2018; McNeish, 2014a), and the main role of any multilevel analysis method is merely to accurately estimate standard errors of regression coefficients. This ensures the overall reliability of CRSE.

3.2 Generalized Estimating Equations

The second common DBM is the generalized estimating equations (GEE; Liang & Zeger, 1986). Its basic steps are: first, the researcher specifies a working correlation matrix structure that reflects the correlation among observations within groups; then, ignoring the multilevel structure (as in single-level regression) and assuming independent errors, a regression model is fitted to obtain regression coefficients and residuals; next, residual information from the previous step is used to estimate initial values of the working correlation matrix to conform to the researcher-specified structure, and the working correlation matrix is used to estimate the covariance matrix of outcome variables (within each group) to update regression coefficients and standard errors to reflect observation dependence; through continuous iteration among the updated working correlation matrix, outcome variable covariance matrix, and regression coefficient estimates until regression coefficients show no obvious change between iterations, the model converges and regression coefficient estimation is complete; finally, the CRSE introduced earlier is used to correct standard errors. For more technical details about GEE, see McNeish (2019).

The main prerequisites for using GEE are: (1) observations from different groups are uncorrelated; (2) the working correlation matrix specification approximates the true structure (McNeish et al., 2017). This method not only can directly analyze level-2 variables like CRSE but also has a unique advantage not found in other DBM: it is suitable for analyzing longitudinal data and categorical outcome variables (McNeish, 2014b; McNeish et al., 2017). This makes it the most comprehensive DBM. A relative limitation of GEE is the lack of available effect size and model fit evaluation indicators, which is not conducive to evaluating effect magnitude and model fit.

3.3 Fixed Effects Model

The fixed effects model (FEM) controls all level-2 variation by including group identifier variables as dummy variables in the regression model (for example, when students are nested within schools, school IDs can be set as dummy variables), thereby preventing the influence of error non-independence when analyzing relationships among level-1 variables (McNeish & Kelley, 2019). Taking students nested within schools as an example, assuming X and Y are student-level independent and dependent variables respectively, and there are 3 schools total, the fixed effects model can be represented as:

$$Y = \beta X + \gamma_0 + \gamma_1 S_1 + \gamma_2 S_2 + r$$

where β is the regression coefficient of Y on X, representing the average effect of the level-1 independent variable on the dependent variable across all schools; S_1 and S_2 are dummy variables representing schools 1 and 2, with school 3 as the reference group; γ_0 represents the mean of the dependent variable for the reference group; γ_1 and γ_2 represent mean differences in the dependent variable between schools 1, 2 and school 3; and r is the residual term, representing only level-1 errors.

The main prerequisite for using FEM is that residuals follow a normal distribution (McNeish & Kelley, 2019). This method' s greatest advantage is that all level-2 variation is completely controlled, bringing two benefits: (1) it completely eliminates estimation bias that may occur in other multilevel analysis methods like HLM due to omission of necessary level-2 covariates (McNeish & Kelley, 2019); (2) for multilevel-structured data, all level-1 independent variables typically require group-mean centering (GMC) to prevent regression coefficients from being inaccurate due to contamination by level-2 effects (Fang et al., 2010). OLS, HLM, CRSE, and GEE all require this step. However, because FEM controls all level-2 variation, reliable level-1 regression coefficients can be obtained without centering, streamlining the procedure. Additionally, FEM has an extremely low application threshold with virtually no statistical software requirements—any software capable of regression analysis can implement FEM.

FEM' s main disadvantages are that, unlike other DBM, it cannot analyze level-2 variables and can only analyze level-1 effects; secondly, FEM treats level-2 sampling units as fixed rather than random, so results cannot be generalized to unsampled level-2 units. However, in single-level research contexts containing multilevel errors, because the number of level-2 groups is too small, even HLM would have difficulty generalizing results to other level-2 units, making this limitation not a major issue.

3.4 Advantages of DBM

3.4.1 Alignment with Single-Level Research Contexts Containing Multilevel Errors When single-level research contains multilevel errors, HLM' s core advantage of analyzing random effects becomes redundant or even burdensome. In contrast, DBM inherently lacks the ability to analyze random effects; its main function is to eliminate the "noise" of multilevel errors in single-level research to ensure accurate standard error estimation for level-1 regression coefficients. This both meets the primary needs of single-level research and avoids wasted functionality, creating better alignment between method and research question.

3.4.2 Fewer Assumptions and Greater Robustness in Parameter Estimation DBM' s inability to specify random effects frees it from reliance on assumptions about random effects, thereby reducing risks associated with violations of such assumptions and providing better robustness. This is an important advantage in single-level research containing multilevel errors, as these assumptions are harder to test and satisfy. For example, McNeish' s (2019) simulation and empirical studies showed that when random slopes are omitted or random slopes are non-normal, GEE maintains accurate parameter estimation, whereas HLM exhibits underestimation of level-1 standard errors. Since CRSE is a step in GEE and is essentially equivalent to GEE in cross-sectional studies with small intraclass correlation coefficients, CRSE can be expected to demonstrate similar robustness under comparable conditions. Additionally, a simulation study by McNeish and Stapleton (2016) found that when level-2 sample size is extremely small, HLM shows low coverage rates for 95% confidence intervals of level-1 regression coefficients and underestimates coefficient standard errors, while FEM performs well. This also demonstrates FEM' s relative robustness.

3.4.3 Fewer Convergence Problems Complex random effects specifications create greater convergence difficulties (McNeish & Stapleton, 2016). DBM' s lack of focus on random effects substantially reduces convergence problems. Numerous simulation studies have shown that convergence rates for CRSE, GEE, and FEM are all higher than for HLM (Bolin et al., 2019; Huang, 2018; McNeish, 2019). Empirical studies also provide evidence. For example, in a single-level mediation study containing multilevel errors by Zhu and Xie (2018), the authors first attempted HLM modeling, which failed to converge, but encountered no problems when switching to CRSE.

3.4.4 Model Parsimony Another prominent advantage of DBM is model parsimony. DBM does not require researchers to consider which level each variable belongs to, nor does it require decomposing model residuals across different levels like HLM. Instead, all level variables and residuals are placed within a single single-level regression model, simplifying the basic model form. For example, for a single-level mediation model containing level-2 errors (i.e., a 1-1-1 mediation model; Fang et al., 2010), the established HLM (assuming no random slopes) is:

$$M_{ij} = \beta_{M0j} + \beta_{M1j}(X_{ij} - \bar{X}_j) + r_{Mij}$$

$$\beta_{M0j} = \gamma_{M00} + u_{M0j}$$

$$\beta_{M1j} = \gamma_{M10}$$

$$Y_{ij} = \beta_{Y0j} + \beta_{Y1j}(X_{ij} - \bar{X}_j) + \beta_{Y2j}(M_{ij} - \bar{M}_j) + r_{Yij}$$

$$\beta_{Y0j} = \gamma_{Y00} + u_{Y0j}$$

$$\beta_{Y1j} = \gamma_{Y10}$$

$$\beta_{Y2j} = \gamma_{Y20}$$

where $\bar{X}_j$ and $\bar{M}_j$ are group means of the independent and mediating variables, representing the between-group components of the two variables; $X_{ij} - \bar{X}_j$ and $M_{ij} - \bar{M}_j$ are group-mean deviations of the independent and mediating variables, representing the within-group components; γ_{M00} and γ_{Y00} are intercept terms for the mediating and dependent variables; γ_{M10} and γ_{M01} are within-group and between-group effects of the front path coefficient a of the mediation effect; γ_{Y20} and γ_{Y02} are within-group and between-group effects of the back path coefficient b of the mediation effect; γ_{Y10} and γ_{Y01} are within-group and between-group components of the direct effect c' of independent variable X_{ij} on Y_{ij} after controlling for the mediating role of M_{ij}; u_{M0j} and r_{Mij} are level-2 and level-1 residual terms for M_{ij}; and u_{Y0j} and r_{Yij} are level-2 and level-1 residual terms for Y_{ij}.

The above system comprises 7 equations with 17 types of coefficients, features nested relationships among equations, and uses symbols with at least double and up to triple subscripts (necessitating at least symbols i and j to represent level-1 individuals and level-2 organizations). Adding random slopes would further complicate the model. However, using CRSE and GEE, the model can be simplified to:

$$M = \gamma_M + a_w(X - X_m) + a_b X_m + r_M$$

$$Y = \gamma_Y + b_w(M - M_m) + b_b M_m + c'_w(X - X_m) + c'_b X_m + r_Y$$

This formula is no different from the basic form of single-level regression. Variables X_j, M_j, $X_{ij} - X_j$, $M_{ij} - M_j$, γ_{M00}, γ_{Y00}, γ_{M10}, γ_{M01}, γ_{Y20}, γ_{Y02}, γ_{Y10}, γ_{Y01}, $u_{M0j} + r_{Mij}$, and $u_{Y0j} + r_{Yij}$ in the HLM equations are replaced by X_m, M_m, $(X - X_m)$, $(M - M_m)$, γ_M, γ_Y, a_w, a_b, b_w, b_b, c'_w, c'_b, r_M, and r_Y, respectively. Because variables and errors are not marked and placed by level, equations are reduced to 2, coefficients to 10 types, and subscripts to at most one level.

If FEM is used (assuming 3 level-2 sampling units), the model can be further simplified:

$$M = \gamma_{M0} + \gamma_{M1} S_1 + \gamma_{M2} S_2 + aX + r_M$$

$$Y = \gamma_{Y0} + \gamma_{Y1}S_1 + \gamma_{Y2}S_2 + bM + c'X + r_Y$$

This model not only eliminates the need for level separation but also, because dummy variables for group identifiers explain all level-2 variation, eliminates the need for group-mean centering of level-1 variables and inclusion of group means in the model. Using raw scores of X and M allows accurate estimation of single-level mediation effects, making it simpler than CRSE and GEE. The mediation models corresponding to several methods are shown in Figure 1 [Figure 1: see original paper], illustrating that models become increasingly parsimonious from left to right.

In summary, DBM is more concise and intuitive than HLM, with model coefficient meanings immediately clear. Researchers familiar with single-level regression can seamlessly transition to DBM, and interpretation of results is the same as for single-level regression.

3.4.5 Operational Simplicity The parsimony of DBM modeling is directly reflected in statistical operations. Taking CRSE as an example, this method is easily implemented in Mplus software, with simpler operation commands than HLM for analyzing the same model. First, HLM' s ANALYSIS command must change according to research levels and random effects of interest, whereas CRSE only requires the fixed commands "estimator=mlr; type=complex." Second, in the VARIABLE and MODEL sections, multilevel linear models must specify commands separately by variable level (a section particularly prone to errors), while CRSE only needs to specify variables and models like single-level regression without level separation. Third, CRSE can directly output model diagrams using the plotting function, whereas HLM cannot. In addition to CRSE, other DBM are also easy to operate; both GEE and FEM can be implemented through window-based operations in SPSS (see Appendix).

3.4.6 Computational Efficiency Because DBM does not analyze random effects, software running speed is significantly faster than HLM. HLM running time heavily depends on random effects, convergence criteria, sample size, and software used. For example, an HLM containing 3 random effects using adaptive Gaussian quadrature for solution requires···

3.5 Summary

In psychological research, DBM, as a relatively new multilevel data analysis method, enables researchers to effectively control multilevel errors and accurately analyze variable relationships without stepping outside the "comfort zone" of single-level research logic. DBM features fewer assumptions, more parsimonious models, simpler operations, more efficient computation, and better robustness, making it a good alternative to HLM in single-level research contexts.

Of course, DBM is not without limitations. The common problem with all DBM is that random effects can only be controlled, not analyzed. For example, if researchers want to understand whether the relationship between teachers' average emotional intelligence and burnout differs across universities, they could analyze this by building a multilevel model with random slopes. Furthermore, researchers could conduct statistical analysis for specific schools, i.e., examine exactly how large the effect of teachers'average emotional intelligence on burnout is for each university. For these questions, HLM is the only available analytical tool; DBM is powerless. DBM can only make a general inference about the average relationship between teachers' emotional intelligence and burnout across all schools, though this is sufficient for single-level research.

4. Application Example

Below, a simulated dataset is used to demonstrate the effectiveness of the DBM introduced in this paper for controlling multilevel errors in single-level research (using a 1-1-1 mediation model as an example) and to compare them with OLS and HLM. According to existing research (Fang et al., 2010), such mediation effects can be decomposed into between-group mediation (higher-level mediation effect, see level-2 portion of model in Figure 1 left panel) and within-group mediation (lower-level mediation effect, see level-1 portion of model in Figure 1 left panel). The latter reflects the mediation effect in the single-level sense, so we focus only on it (for simplicity, the front and back path coefficients of within-group mediation are directly denoted as a and b). The true values of coefficients a and b are both 0.30, with true standard errors of 0.029 and 0.041, respectively; the true value of the mediation effect ab is 0.09, with a true standard error of 0.016. To better simulate a single-level research context with multilevel errors (small level-2 sample size, large level-1 sample size), our generated data have only 3 groups at level-2, with 200 individuals per group, for a total level-1 sample size of 600, approximating typical cluster sampling effects. The ICCs of the dependent variable Y and mediating variable M are 0.96 and 0.30, respectively, indicating a clear multilevel structure.

The data were analyzed using OLS, OLS with group-mean centering of predictor variables (OLS-GMC), HLM, CRSE, GEE, and FEM (operation templates are provided in the Appendix). All methods except OLS and FEM pre-centered predictor variables using group-mean centering. Results (see Table 2) show that except for the front path of mediation, OLS overestimates all other coefficients and standard errors in the mediation model. After group-mean centering predictor variables, OLS bias in regression coefficient estimation is completely eliminated, but standard errors for coefficient b and ab remain substantially overestimated. These results demonstrate that OLS cannot control multilevel errors without any standard error correction.

For all DBM and HLM, regression coefficient estimates are accurate. All three DBM show small biases in estimating all standard errors (except that CRSE' s relative estimation bias for ab standard error is slightly above 10%), while HLM underestimates all standard errors. Overall, FEM shows the best performance, followed by GEE, then CRSE, with HLM performing relatively worst.

In terms of operational process, FEM is simplest, enabling not only window-based operation through SPSS but also eliminating the need for centering and controlling between-group variables like other methods. Using raw scores of X and M and dummy variables for group identifiers is sufficient. Furthermore, with the popular SPSS mediation analysis macro PROCESS, FEM can simultaneously implement both the causal steps approach and coefficient product approach with bootstrap intervals, and can complete window-based operations for many complex mediation and moderation mixture models (HLM cannot provide bootstrap intervals). GEE can also be operated through SPSS windows, though it cannot directly test the coefficient product. Both CRSE and HLM require writing Mplus syntax, but CRSE does not involve level separation operations (i.e., commands related to between and within), making it simpler and less error-prone than HLM.

5. Method Selection Recommendations

The above introduces methods that can control multilevel errors in single-level research. How should researchers choose among them in applied studies? Based on researchers' actual research purposes and the characteristics of each method, the following recommendations are offered:

(1) Considering that DBM better matches single-level research contexts, when researchers lack sufficient HLM knowledge and have no need to analyze random effects, DBM is recommended as the priority choice, particularly FEM, given its highest parameter estimation accuracy, simplest operation, and easy integration with PROCESS for window-based analysis of various complex models.

(2) If researchers want to analyze some level-2 fixed effects in addition to level-1 effects, such as contextual effects of level-1 independent variables, FEM is inadequate. In this case, GEE or CRSE is recommended, but the level-2 sample size should not be too small when using these methods (Huang, 2018; McNeish & Stapleton, 2016).

(3) When researchers have sufficient HLM background and need to examine random effects, they should deliberately collect multilevel data, particularly ensuring adequate level-2 sample size. If the model can converge, researchers are advised to use HLM incorporating all possible random effects with robust maximum likelihood estimation (MLR) as the parameter estimation method, which can minimize problems caused by random effects assumption violations and enable analysis of random effects.

(4) Both HLM and DBM require group identifier variables for data analysis. Even when conducting single-level research, it is recommended to retain group identifier information for each sampling unit when collecting data,

to prevent the actual data level from exceeding the expected level, which would result in failure to control multilevel errors and reduce research result reliability.

Acknowledgments

Thanks to Associate Professor Wang Huihui from the School of Education at Ningxia University and Dr. Dai Buyun from the School of Psychology at Jiangxi Normal University for their assistance in writing this paper.

References

Fang, J., Zhang, M. Q., & Qiu, H. Z. (2010). Multilevel mediation effect based on hierarchical linear theory. *Advances in Psychological Science*, 18(8), 1329-1338.

Zhu, Y., & Xie, B. B. (2018). The relationship between perceived differential climate and employee silence: The mediating role of affective commitment and the moderating role of traditionality. *Acta Psychologica Sinica*, 50(5), 539-548.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Retrieved May 26, 2018, from https://arxiv.org/pdf/1506.04967.pdf

Bolin, J. H., Finch, W. H., & Stenger, R. (2019). Estimation of random coefficient multilevel models in the context of small numbers of level 2 clusters. *Educational and Psychological Measurement*, 79(2), 217-248.

Cameron, A. C., & Miller, D. L. (2015). A practitioner' s guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372.

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education*, 84(1), 175-196.

Huang, F. L. (2018). Multilevel modeling and ordinary least squares regression: How comparable are they? *Journal of Experimental Education*, 86(2), 265-281.

Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J. M., & Thiébaut, R. (2007). Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics and Data Analysis*, 51(10), 5142-5154.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22.

Litière, S., Alonso, A., & Molenberghs, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics*, 63(4), 1038-1044.

McNeish, D. M. (2014a). Analyzing clustered data with OLS regression: The effect of a hierarchical data structure. *Multiple Linear Regression Viewpoints*, 40(1), 11-16.

McNeish, D. M. (2014b). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological Methods*, 19(4), 552-579.

McNeish, D. M. (2019). Effect partitioning in cross-sectionally clustered data without multilevel models. *Multivariate Behavioral Research*, 54(6), 862-877.

McNeish, D. M., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, 24(1), 20-35.

McNeish, D. M., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, 51(4), 495-518.

McNeish, D. M., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114-140.

McNeish, D. M., & Wentzel, K. R. (2017). Accommodating small sample sizes in three-level models when the third level is incidental. *Multivariate Behavioral Research*, 52(2), 200-215.

Appendix: Software Operations for Controlling Multilevel Errors in Single-Level Mediation Analysis

HLM Mplus Syntax

```
DATA: FILE = 示例数据.csv;
VARIABLE: NAMES = id y group x m xc mc xm mm;
          !group = group identifier variable; xc, mc = group-mean deviations of independent
          Usevariables= y group xc mc xm mm;
          CLUSTER=group; !Specify group identifier variable
          WITHIN = xc mc; !Specify level-1 variables
          BETWEEN = xm mm; !Specify level-2 variables
ANALYSIS: estimator=MLR; TYPE = TWOLEVEL;
MODEL: %WITHIN%
         y on xc mc;
         mc on xc; !Specify level-1 model
       %BETWEEN%
         y on xm mm;
         mm on xm; !Specify level-2 model
MODEL indirect: y ind xc; !Analyze within-group component of mediation model
OUTPUT: STANDARDIZED(STDYX);
```

CRSE Mplus Syntax

```
DATA: FILE = 示例数据.csv;
VARIABLE: NAMES = id y group x m xc mc xm mm;
          Usevariables= y group xc mc xm mm;
          CLUSTER=group;
ANALYSIS: estimator=MLR; TYPE=COMPLEX; !Key commands for using CRSE
MODEL: y on xm mm xc mc;
       mc on xc;
       mm on xm; !Specify model without level separation
MODEL indirect: y ind xc;
OUTPUT: STANDARDIZED(STDYX);
```

FEM SPSS Window Operation Steps

1. In the SPSS menu bar, select: Transform → Create Dummy Variables, and move the group identifier variable "group" into the input box on the right. Then enter the root name for dummy variables (use the original variable name "group") and click "OK." After generating the group identifier dummy variables, arbitrarily delete one column.
2. In the SPSS menu bar, select: Analyze → Regression → PROCESS (assuming PROCESS plugin is already installed), move independent variable x, dependent variable y, and mediating variable m into their respective input boxes, and move the generated group identifier dummy variables into the Covariates box. Select Model Number 4 (representing simple mediation model), then click "OK."

GEE SPSS Window Operation Steps

1. In the SPSS menu bar, select: Analyze → Generalized Linear Models → Generalized Estimating Equations.
2. In the "Repeated"tab: Move the grouping variable "group"into the "Subject variables" box, and select "Model-based estimator" in the "Covariance Matrix" radio button.
3. In the "Response" tab: Specify the outcome variable. When analyzing the front path of mediation, move the mediating variable m into the box; when analyzing the back path, move the dependent variable y into the box.
4. In the "Predictors" tab: Specify predictor variables. When analyzing the front path, move independent variables (including deviation xc and group mean xm, 2 variables total) into the "Covariates" box; when analyzing the back path, move both independent and mediating variables (including variable deviations xc, mc and group means xm, mm, 4 variables total) into the "Covariates" box.
5. In the "Model" tab: Move all predictor variables into the "Model" box, then click "OK."

Note: Because this is the causal steps approach, the front and back paths of mediation must be operated separately.

Note: Figure translations are in progress. See original paper for figures.