Understanding HLM Models and Type VI Errors: The Need for Reflection

Susan M. Tracz — California State University, Fresno
Isadore Newman — Florida International University
David O. Newman — Florida Atlantic University

This article presents a framework for two common hierarchical linear models (HLM), instructions to run them in Statistical Package for the Social Sciences (SPSS), and a comparison between SPSS 21.0 linear mixed models (LMM) and HLM 7.0 output. Discussions topics include centering in hierarchical modeling, a comparison of SPSS output for the default restricted maximum likelihood and maximum likelihood solutions, a comparison of SPSS output for HLM and ordinary least squares (OLS) multiple linear regression (MLR) with person vectors output on mean square errors and R^2, and a comparison of R^2 and R^2 changes. Correlated residuals between SPSS LMM and OLS MLR provide a context for considering hypothesis testing, research questions, and the choice of statistical tests. Finally, this article addresses the complexity of developing multi-level linear research questions and determining which statistical techniques are appropriate for answering those questions so Type VI errors can be avoided.

While much research in education and the social sciences is conducted on individuals, in many cases, individuals exist within larger contextual systems. For example, Jessor (1993) discusses adolescents living within the family, the school, and the neighborhood, and each of these systems exerts influences on individuals. Additional systems are numerous and varied, and other examples include students within classrooms, classrooms within schools, patients within hospitals, workers within employment settings, to name a few (Bickel, 2007; Newman & Newman, 2012; Radenbush & Bryk, 2002). Situating an individual within a larger context or group is called "nesting." Nesting affects outcomes in unique ways, and the variation due to nesting must be incorporated into statistical models used to examine outcomes.

When recognized, nesting had been previously addressed with disaggregation and aggregation, which are various methods for assigning the same value to all individuals within a group, for example, the same value on teacher self-efficacy to all students within a class. This approach violates the assumption of independence, inflates the Type I error rate, and results in correlated residuals, which generally lead to lower standard errors, higher probabilities of rejecting null hypotheses, and inflated R^2s (Bickel, 2007). Also, results were often interpreted on the individual level even though group level data were analyzed. This is referred to as an ecological fallacy or the Robinson (1950) effect, which is a logical fallacy that occurs when statistical outcomes for groups are attributed to specific individuals of that group, under the assumption that all members of the group are identical or have the same attributes.

As can be seen, when data are nested and are not analyzed appropriately, numerous problems arise. Most prominent among these is the violation of the assumption of independence. Independence can be defined in three ways. First, in repeated measures analyses, the multiple measures, or scores for individuals, are not independent of each other, but are correlated to each other (Cohen & Cohen, 1983: Pedhazer, 1982). Second, independent variables within statistical analyses are often related or correlated to each other, which is called multicollinearity (Cohen & Cohen; Pedhazer). Finally, independence is related to the sampling of individuals from populations for participation in research studies, which utilize statistical analyses that have assumptions that must be adhered to (McNeil, Newman, & Frass, 2012; McNeil, Newman & Kelly, 1996). Each individual in a population who is sampled for participation in a study should have an equal probability of being chosen. However, when those individuals are chosen in groups or clusters, such as members within families or students within classrooms, the assumption of independence is violated. This is the type of non-independence that is the focus of this article.

When data are nested and are not analyzed appropriately, an even more basic problem can occur. Researchers often do not know what research questions they are asking or do not know that they are not using the appropriate statistical analyses to answer their questions (Clason & Mundfrom, 2012; Nimon & Henson, 2010; Tracz, Nelson, Newman & Beltran, 2005). In these cases, misinterpretations are likely or even inevitable. The problem of utilizing models or statistical analyses, which do not address the research questions of interest, has been labeled a "Type VI error" (Newman, Frass, Newman, & Brown, 2002). It is obvious that meaningful interpretations will be precluded in these cases.

In his aptly titled book, *Multilevel Analysis for Applied Research: It's Just Regression,* Bickel (2007) notes that the

> development of regression analysis over the last 100 years consists largely of efforts to find ways to make OLS multiple regression applicable to increasingly complex questions, even when its basic assumptions have been violated . . . Moreover, whatever adjustments are made to accommodate violations of well-known regression assumptions, analysts know that they are still doing regression analysis. (p. 4)

Hierarchical linear modeling (HLM) is one statistical advance created to rectify the problems associated with nested data. Radenbush and Bryk (2002) use the term HLM to describe models used for data having specific nested structures. Their work builds on that of Lindley and Smith (1972) and Smith (1973) who proposed linear models using Bayesian estimation methods for nested data with complex patterns of errors. Difficulties in estimating covariance components with unbalanced data were addressed by Dempster, Laird, and Rubin (1977) who developed an approach for covariance component estimation that was more widely applicable. Additional methods, including hierarchical linear modeling (Burk, Raudenbush, Seltzer, & Congdon, 1988), were possible due to these advances.

HLM models are referred to as multi-level linear models, mixed-effects models, random-effects models, and covariance components models in various disciplines (Radenbush & Bryk, 2002). Many researchers (Field, 2009; Kreft, 1996; Morris, 1995; Mundfrom & Schultz, 2002; Raudenbush & Bryk; Tabachnick & Fidell, 2007) believe that HLM is superior to ordinary least squares (OLS) multiple linear regression (MLR). Newman and Newman (2012) define HLM as "the general case of *multiple linear regression* [italics added] and a preferred statistical method for analyzing complex nested data structures" (p. 1); however, the HLM terminology does not parallel regression terminology. These differences in the conceptualization of research questions and the use of statistical terminology can be confusing to students in graduate programs in the U.S. who take initial statistics courses, which cover correlation, regression, *t*-tests, analysis of variance (ANOVA), and chi-square, and that teach classical hypothesis testing and the formulation of classical research questions. Many have difficulty jumping to advanced HLM models and generally consider them to be a completely different class of analyses (Mundfrom & Schults, 2001, 2002). In fact, "failure to draw all but the most abstract parallels between multi-level analysis and conventional regression analysis" fosters this confusion (Bickel, 2007, p. 4), and many do not realize that HLM is a type of regression. Not understanding the relationship between regression models and hierarchical linear models also contributes to difficulty in formulating appropriate research questions. Many researchers retain their classical training in developing research questions; though, they are using HLM.

Confusion about what research questions are addressed with HLM and the resulting Type VI errors are widespread (Newman et al., 2002). This uncertainty is compounded as the straightforward terminology used with classical regression, such as estimating beta weights; mean differences; and interaction, is substituted with new language and terms including means as outcomes, intercepts as outcomes, slopes as outcomes, and centering (Newman, Newman, & Salzman, 2010).

Despite the generation of new statistical methods to deal with violations of assumptions with OLS/MLR techniques (Bickel, 2007), in practice, the general populace typically uses the same questions that fall into ANOVA-type questions about group differences, regression questions about continuous IVs, and questions of interaction. This is true whether analyses are conducted between categorical and categorical variables, categorical and continuous variables, or continuous and continuous variables (Newman & Newman, 2012). However, classical OLS questions may not be the same questions as those answered with HLM models. In fact, the danger of making a Type VI error (Newman, et al., 2010) may be great.

Purpose

The purposes of this article are to present a framework of HLM models with instructions in Statistical Package for the Social Sciences (SPSS) and a comparison HLM 7.0 and SPSS 21.0 linear mixed models (LMM) output. Information on centering in hierarchical modeling, a comparison of the default restricted maximum likelihood (REML) and maximum likelihood (ML) solutions in SPSS, a comparison of SPSS HLM and OLS MLR with person vectors output on mean square errors and R^2, R^2 and R^2 changes, and correlated residuals between SPSS LMM and OLS MLR are also discussed. Finally, this paper addresses the complexity of developing multi-level linear research questions and determining which statistical techniques are appropriate for answering those questions so that Type VI errors can be avoided.

Data Set

This article uses the High School and Beyond Data set that was collected to examine the effects of socioeconomic status, school and school type, and other variables on high school student math achievement (Bryk, Holland, Lee, & Carriedo, 1984). The variables selected here for use in the HLM and SPSS analyses are defined below. Examples of values for these variables appear in Table 1.

Stud ID = A consecutive number assigned to each student within each school.

SCHOOL = A unique number assigned to each school.

MATHACH = Individual math achievement score for each student.

SES = Individual socio-economic status for each student.

SCH1 – SCH160 = Dummy coded vectors for each school which act as person vectors.

SES_SCH1 – SES_SCH160 = Individual SES multiplied by the dummy coded vectors for each school.

Models

Two sets of HLM and OLS/MLR models will be used in this study. They include a null model, which is used to calculate the interclass correlation coefficient (ICC), and a second model that has one variable, SES, added to level one. The formula for the ICC, which is the proportion of total variance in the dependent variable that is attributable to the context or nested variable (Field, 2013), is:

$$ICC = \tau_{00}/(\tau_{00} + \sigma^2)$$

For these data, the ICC is .82, which indicates dependency among the scores and a strong influence of the nested school variable. These models in HLM and MLR, along with the research questions, are given below. Instructions to run HLM in SPSS 21.0 appear in Appendix A, SPSS 21.0 output appears in Appendix B, and the HLM Version 7.0 output appears in Appendix C.

<u>Model 1</u>:

HLM	Level 1:	$MATHACH = \beta_{0j} + r_{ij}$
	Level 2:	$\beta_{0j} = \gamma_{00} + u_{0j}$
	Combined:	$MATHACH = \gamma_{00} + u_{0j} + r_{ij}$
OLS/MLR	Full Model:	Math = a0U + b1SCH1 +. . . + bnSCHn + e
	Restricted Model:	Math = a0U + e

Research Question: Are there school differences in math achievement? This is a simple, straight forward research questions with a single model that aligns with classical hypothesis testing.

<u>Model 2</u>:

HLM	Level 1:	$Math = \beta_{0j} + \beta_{1j}SES + r_{ij}$	
	Level 2:	$\beta_{0j} = \gamma_{00} + u_{0j}$	
		$\beta_{1j} = \gamma_{10} + u_{1j}$	(SES is group centered)
	Combined :	$Math = \gamma_{00} + \gamma_{10}\,SES + u_{0j} + u_{ij}SES + r_{ij}$	
OLS/MLR			
	Full Model:	Math = a0U + b1SCH1 +. . . + bnSCHn + c1SES_SCH1 + SES_SCHn + e	
	Restricted Model:	Math = a0U + b1SCH1 +. . . + bnSCHn + e	

Research Questions: The research question for the OLS regression test is: Does the school * SES interaction account for a significant proportion of unique variance in math achievement over and above the main effects of SES and school? This is the classical definition of interaction (Hayes & Matthers, 2009). However, HLM is testing whether there is a significant main effect for SES, and whether there are significant slope differences for SES by schools. The HLM test of multiplicative variables representing slope differences is not interaction in the classical sense since it does not test for variance over and above the variance accounted for by the main effects. In the literature, this is a common Type VI Error that occurs when people use HLM. It is possible to test the same classical research question as OLS regression, but this can only be done with sets of models that are further explained later in this paper.

Table 1. Data Examples for the Variables for this Study

StudID	SCH	MATHACH	SES	SCH1	SCH 2	SES_SCH1	SES_SCH2
1	1	6.5	-.12	1	0	-.12	0
2	1	9.2	.14	1	0	.14	0
3	1	4.5	-.12	1	0	-.12	0
4	1	11.9	1.12	1	0	1.12	0
5	1	16.8	.02	1	0	.02	0
1	2	23.1	.78	0	1	0	.78
2	2	11.5	.32	0	1	0	.32
3	2	7.5	-.25	0	1	0	-.25
4	2	13.2	-.98	0	1	0	-.98
5	2	5.5	.48	0	1	0	.48
6	2	5.8	.22	0	1	0	.22

Table 2. Model 1 - One-way Random Effects ANOVA Model using REML Using SPSS and HLM

Fixed Effects		Coefficients (SE)	*t*(df)	*p*
Model for mean school math achievement (β_0)				
Intercept (γ_{00})	SPSS	12.64 (.24)	51.71 (156.65)	<.001
	HLM	12.64 (.24)	51.87 (159)	<.001
Random Effects (Variance Components)		**Variance**	**Wald χ^2(df)**	***p***
Variance in school means (τ_{00})	SPSS	8.61	59.26	<.001
	HLM	8.61	1660.23 (159)	<.001
Variance within schools (σ^2)	SPSS	39.15		
	HLM	39.15		

Table 3. *Model 2 – Random Coefficients Model with REML (Group-mean centering) between SPSS and HLM Output*

Fixed Effects		Coefficients (SE)	*t*(df)	*p*
Model for mean school math achievement (β_0)				
Intercept (γ_{00})	SPSS	12.64 (.24)	51.68 (156.75)	<.001
	HLM	12.66 (.19)	66.92 (159)	<.001
Model for SES slope (β_1)				
Intercept (γ_{10})	SPSS	2.19 (.13)	17.10 (155.22)	<.001
	HLM	2.39 (.11)	20.34 (159)	<.008
Random Effects (Variance Components)		**Variance**	**Wald χ^2(df)**	***p***
Variance in school means (τ_{00})	SPSS		8.04	<.001
	HLM		905.26 (159)	<.001
Variance in SES slopes (τ_{11})	SPSS	.69	2.47	.03
	HLM	.65	216.21 (159)	.002
Variance within schools (σ^2)	SPSS	36.70		
	HLM	36.83		

Comparison of HLM and SPSS Output for Models 1 and 2

Tables 2 and 3 present the output using REML estimation for models 1 and 2; employing the Radenbusch and Bryk (2002) HLM program (v. 7.0) and SPSS (v. 21.0). For model 2, the random coefficients model with REML is also called the unconditional model and has no level 2 predictors. As can be seen, the output for the intercept (γ_{00}), variance in school means (τ_{00}), variance within schools (σ^2), *t* and *p* values for models 1 and 2 are nearly identical. Only the test of variance in school means (τ_{00}), where SPSS reports the Wald statistic and HLM reports a χ^2, are different; though, the *p* values are the same.

Centering

Many, if not most, linear mixed models use centering of predictors in level-1 and level-2. Centering is done by subtracting the mean from each score so that the distribution of the resulting difference scores is centered on zero and those centered scores represent the distance above or below the average score. However, the decision to center should not be done by default, but instead should be based upon the researcher's specific questions of interest (Newman & Newman, 2012), which again reinforces the concept that Type VI errors should be avoided. There are two types of centering that should be aligned with two decisions that need to be made. The first type is group mean centering where each score is centered in relationship to the higher-order structure. In this study, one option would be to group mean center with each of the students' SES scores centered within one of the 160 schools they attend. The second type is grand mean centering, which would center the students' SES in relation to the average SES of all of the students within all of the schools. So the two decisions are: 1). should centering be done? and 2). if yes, should group mean or grand mean centering be done?

Many people prefer grand mean centering over group mean centering (Burton, 1993; Hoffman & Gavin, 1998; Kreft, de Leeuw, & Aiken, 1995). Sarkisian (2007) suggests the original scores should not be used if zero is not a meaningful value. Despite these findings, Field (2009, 2013) suggests that centering is not an easy issue to decide and requires that the researcher have a solid understanding of the data and the analyses being conducted. Field does suggest that centering is a useful method of reducing the problem of multicollinearity between independent variables, especially if the predictor does not have an interpretable zero. Therefore, if the question of interest is what variance is accounted for by the relative position of that subject within the group, then group mean centering should be used. However, if the researcher is interested in the absolute value of the predictor, then grand mean centering should be used. If grand mean centering is used, then the intercepts become adjusted grand means and have no effects on the slopes. On the other hand, if group mean centering is used, the intercepts are the means of each group and may change the meaning of the coefficients, making them difficult to interpret since the values obtained vary across groups. It must be remembered that if one uses group mean centering, it is only the person level variables that are of interest. It is also important to remember that if one does decide to use group mean centering instead of grand mean centering, the group centered variable should be assigned at level-2 as long as one is not interested in testing the unique variance accounted for by the group effects.

As stated earlier, this complicated decision of whether to center using groups or grand mean centering is not a statistical issue, but one that is determined by the research questions of interest in each particular study. Failing to align specific questions of interest to the centering choice many very well lead to a Type VI error (Newman & Newman 2012).

Comparisons of REML and ML SPSS Output for Model 2

A comparison of the REML and ML estimates in SPSS are exactly the same for some parameters including as the intercept, γ_{00}, which is the mean for all schools on math achievement, for the SES slope (β_1), and for the variance within schools (σ^2). There are minor differences in the results for the variances

Table 4. Model 2 – Random Coefficients Model using SPSS with REML and with ML

Fixed Effects		Coefficients (SE)	*t*(df)	*p*
Model for mean school math achievement (β_0)				
Intercept (γ_{00})	REML	12.64 (.24)	51.68 (156.75)	<.001
	ML	12.64 (.24)	51.85 (157.73)	<.001
Model for SES slope (β_1)				
Intercept (γ_{10})	REML	2.19 (.13)	17.10 (155.22)	<.001
	ML	2.19 (.13)	17.15 (156.15)	<.001
Random Effects (Variance Components)		Variance	Wald	*p*
Variance in school means (τ_{00})	REML	8.68	8.04	<.001
	ML	8.62	8.06	<.001
Variance in SES slopes (τ_{11})	REML	.69	2.47	.03
	ML	.68	2.44	.02
Variance within schools (σ^2)	REML	36.70		
	ML	36.70		

Table 5. Model 1 SPSS MLR Results

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.437[a]	.191	.173	6.2563275

a. Predictors: (Constant), School_160, . . ., School_1
b. Dependent Variable: mathach

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	64906.957	159	408.220	10.429	.000[b]
	Residual	274969.977	7025	39.142		
	Total	339876.934	7184			

a. Dependent Variable: mathach
b. Predictors: (Constant), School_160, . . . , School_81

Table 6. Model 2 SPSS MLR Results

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.437[a]	.191	.173	6.2563275	.191	10.429	159	7025	.000
2	.508[b]	.258	.224	6.0597238	.067	3.895	160	6865	.000

a. Predictors: (Constant), School_160, . . ., School_1
b. Predictors: (Constant), School_160, . . ., SES_School_87
c. Dependent Variable: mathach

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	64906.957	159	408.220	10.429	.000[b]
	Residual	274969.977	7025	39.142		
	Total	339876.934	7184			
2	Regression	87792.398	319	275.211	7.495	.000[c]
	Residual	252084.537	6865	36.720		
	Total	339876.934	7184			

a. Dependent Variable: mathach
b. Predictors: (Constant), School_160, . . ., School_81
c. Predictors: (Constant), School_160,. . . , SES_School_87

for school means (τ_{00}) and SES slopes (τ_{11}) with the ML solution giving a slightly smaller value. For all of the *t*-values, their degrees of freedom and the Wald statistics, except for the Wald for variance in school means, the ML solution again provides smaller values. As Table 4 indicates, these differences are quite small.

Comparison of SPSS HLM and SPSS OLS MLR with Person Vector Models

One way models can be compared is to see if those models produce the same results. This section compares the results for the first model, using HLM generated from SPSS and OLS multiple linear regression, with person vectors using SPSS. The results will consist of the mean square residuals and the R^2s. The correlations between the residuals between the models will also be reported. The SPSS HLM results have already been provided in Tables 2 and 3, and the OLS MLR SPSS for models 1 and 2 are provided above in Tables 5 and 6, respectively.

The mean square residual from the MLR Model 1 results of 39.142 in Table 5 is nearly identical to the variance within schools, 39.15, presented in Table 2. Again for model 2, the mean square residual from the MLR results of 36.720 in Table 6 is nearly identical to the variance within schools of 36.70 presented in Table 3. These values summarized in Table 7 are essentially equal; thus, demonstrating with person vectors how very comparable are HLM and OLS MLR.

R^2 and R^2 Changes

In regression, the R^2 and R^2 changes are easy ways for researchers to quickly assess the overall variance accounted for in different models. Unlike traditional OLS regression, HLM does not produce an R^2. However, by using the following formulas, it is possible to calculate an R^2 for each of the two levels in an HLM model:

Level 1: R^2 at Level 1= $1\text{-}(\sigma^2_{cond}\text{-}\tau_{cond})/\ \text{-}(\sigma^2_{uncond}\text{-}\tau_{uncond})$

Level 2: R^2 at Level 2= $1\text{-}[(\sigma^2_{cond}/n_h)+\tau_{cond}]\ /\ [(\sigma^2_{uncond}/n_h)+\tau_{unccond}]$

Where n_h = the harmonic mean of *n* for the level 2 units

Given these formulae, the question shifts from whether the R^2 can be calculated, to whether it should be calculated? Snijders and Bosker (1994) and Recchia (2010) found that the individual R^2 obtained from each level is often misleading, and as he stated, it "does not behave as one would expect" (p. 3). There is a simple reason for this. The reason is that the variance accounted for at the first level is being moderated or mediated by the Level-2 variables. This concept is similar to the possibility of making a Type VI error when using analysis of covariance (Nimon & Henson, 2010; Tracz et al., 2005). Since the total variance in the dependent variable is being reduced by the covariates, the construct might change, and in all likelihood, will change. In this case, the coefficients of the level-1 predictors, and therefore, the variance they account for, are being modified by the level-2 variables. Recchia and others suggest only looking at R^2s for the combined model. The following equation can be used to obtain the R^2 of the combined linear mixed model.

Combined Model: $R^2= (\sigma^2_{baseling}\text{-}\tau_{cond})/\ \tau_{baseline}$

Hypotheses Testing, Research Questions and Choices of Statistical Tests

OLR regression and most other GLM tests of significance are designed to answer specific research questions. Usually these tests of significance address the findings of the research question within a single analysis. Sometimes, this initial analysis is conducted in hierarchical blocks, but it is still conducted within a single analysis. In contrast, with HLM, this process is conducted in sets of analyses. Raudenbush and Bryk (2002) and Field (2009, 2013) discuss hypotheses testing using LMM. In LMM one has to hand calculate the differences statistic from the baseline or unconditioned model and the conditioned model. Field refers to this as conducting a χ^2 difference test using the 2-log likelihood estimates and the number of parameters. Using the HLM software, Raudenbush and Bryk suggest using the covariance deviance component and the number of estimated parameters in the first model and then test the χ^2 change with the deviance component and the number of parameters in the second model. This χ^2 difference test can then be converted into a R^2 change that can be compared to the R^2 change in the OLS regression or LMM. To convert the χ^2 to an R^2, one can use an effect size converter or the following formula:

$$R^2 = (\chi^2/(\chi^2+N))^{1/2}.$$

When doing hypothesis testing with the HLM software or the SPSS LMM, one has to remember to change the default estimation settings from REML to ML estimates. This is because the χ^2 difference test is comparing the variance accounted for with the fixed effects models. REML is best suited for testing the random effects parameters while ML should be used when testing the fixed effects parameters (Field, 2013; Raudenbush & Bryk, 2002). Even though our comparisons between the REML and ML variance estimates produced similar results, it is suggested that ML are still used for hypothesis testing.

Six models were run in SPSS for this study: the HLM null models using REML and ML, the HLM models with SES using REML and ML, the null model with person vectors in MLR, and the MLR model with SES and person vectors. Residuals were computed for each of these six models and the correlation matrix is presented in Table 8. The correlation coefficients between the residuals for Model 1 using REML and ML and for Model 2 using REML and ML are both 1.00. This is further information supporting the similarities reported above. SPSS uses REML as the default, but the perfect correlation between the errors for these two methods provides evidence that these are virtually identical techniques.

Table 7. Comparison of Mean Square Residual and R^2 Results for SPSS HLM and SPSS OLS MLR for Models 1 and 2

Statistic and Model of Interest	SPSS HLM Results	SPSS OLS MLR Results
Mean Square Residual		
Model 1	39.15	39.142
Model 2	36.70	36.720
R^2		
Model 1	*	*
Model 2	.0665	.067

Note. * The R^2 at Level-1 is being moderated or mediated by the Level-2 variables and can be misleading so it is not calculated or presented here. The R^2 for Level-2 was calculated using Hypothesis Testing in HLM 7.0. The basic formula is reported in the next section.

Table 8. *Correlations Among Residuals for Varying HLM and MLR Models*

Residuals	1	2	3	4	5	6
1) Model 1: HLM/ REML	-					
2) Model 1: HLM/ ML	1.000**	-				
3) Model 2: HLM/ REML	.971**	.971**	-			
4) Model 2: HLM/ML	.971**	.971**	1.000**	-		
5) Model 1: MLR	.999**	.999**	.970**	.970**	-	
6) Model 2: MLR	.956**	.956**	.991**	.991**	.957**	-

Note. For all Correlation N = 7185 and $p < .001$.

Further support that HLM and MLR with person vectors provide the same results is found when examining the correlations among these residuals. The correlation between the HLM – REML residuals and the MLR residuals for model 1 is .999, and the correlation between the HLM – ML residuals and the MLR residuals for model 1 is also .999. The correlation between the HLM – REML residuals and the MLR residuals for model 2 is .991, and the correlation between the HLM – ML residuals and the MLR residuals for model 1 is also .991. This is, again, support that these two statistical methods are providing the same results in response to the same questions. These results are similar to those of Mundfrom and Schults (2002).

Conclusions

This study compared multi-level linear models with HLM and SPSS output and found them to be very similar; though, there appear to be some differences in algorithms or rounding because the values are not always exact. A comparison of the default REML and ML solutions in SPSS found these results to be very similar as well, and the errors for these two methods were perfectly correlated. A comparison of SPSS HLM and MLR with person vectors also found the output to be very similar on mean square residuals, R^2s, and correlated errors. If the errors between different models are correlated in excess of .99, it is reasonable to conclude that those models are addressing the same research questions with the same variables. What seems to be different between HLM and MLR with person vectors is not only the terminology and vocabulary of the research questions, but the basic process of conducting research. In the classical hypothesis testing procedures for regression and ANOVAs, a research questions was asked, a model developed, a statistical analysis performed, and the questions were considered answered either in the affirmative or not. However, the HLM process seems to be more fluid with variables being added and deleted until significance is found and R^2s are reported for models instead of variables. At any rate, researchers need to exert care so that their statistical analyses match their research questions and they avoid making Type VI errors.

References

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!* New York: Guilford.

Bryk, A. S., Holland, P. B., Lee, B. E., & Carriedo, R. A. (1984). *Effective catholic schools: An exploration.* Washington, DC: National Catholic Education Association.

Burk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, R. (1988). *An introduction to HLM: Computer program and user's guide* (2nd ed.). Chicago: University of Chicago Department of Education.

Burton, B. (1993). *Some observations on the effect of centering on the results obtained from hierarchical linear modeling*. Washington, DC: National Center for Education Statistics, U. S. Department of Education.

Clason, D. L., & Mundfrom, D. J. (2012). Adjust means in analysis of covariance: Are they meaningful? *Multiple Linear Regression Viewpoints, 38*, 16-25.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1-8.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: Sage.

Field, A. (2013). *Discovering statistics using IBM SPSS* (4th ed.). Thousand Oaks, CA: Sage.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions In OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods, 41*, 924–936.

Hoffman, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research organizations. *Journal of Management, 24, 623- 641.*

Jessor, *R. (*1993). Successful adolescent development among youth in high-risk settings. *American Psychologist, 48,* 117-126.

Kreft, G. G. (1996). *Are multi-level techniques necessary? An overview, including simulation studies.* Retrieved from http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html.

Kreft, I. G. G., De Leeuw, J., & Aiken, L. S. 1995. The effect of different forms of centering in Hierarchical Linear Models. *Multivariate Behavioral Research, 30*, 1-21.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B, 34,* 1-41.

McNeil, K., Newman, I., & Frass, J. W. (2012). *Designing general linear models to test research hypotheses.* Lanham, MD: University of America Press.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

Morris, C. (1995). Hierarchical models for educational data: An overview. *Journal of Educational and Behavior Statistics, 20,* 190-200.

Mundfrom, D. J., & Schults, M. R. (2001). A comparison between hierarchical linear modeling and multiple linear regression in selected data sets. *Multiple Linear Regression Viewpoints, 28,* 3-11.

Mundfrom, D. J., & Schults, M. R. (2002). A Monte Carlo simulation comparing parameter estimates from multiple linear regression and hierarchical linear modeling. *Multiple Linear Regression Viewpoints, 28,* 18-21.

Newman, I., Frass, J., Newman, C., & Brown. R. (2002). Research practices that produce Type VI errors. *Journal of Research in Education, 12*, 138-145.

Newman, D., & Newman, I. (2012). Multilevel modeling: Clarifying issues of concern. *Multiple Linear Regression Viewpoints, 38*, 26-33.

Newman, D., Newman, I., & Salzman, J. (2010). Comparing OLS and HLM models and the questions they answer: Potential concerns for Type VI errors. *Multiple Linear Regression Viewpoints, 36*, 1-8.

Nimon, K., & Henson, R. K. (2010, May). *Validity of residualized dependent variables after pretest covariance corrections: Still the same variable?* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart and Winston.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Recchia, A. (2010) R-squared measures for two-level hierarchical linear models using SAS. *Journal of Statistical Software, 32*, 1-9.

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

Sarkisian, N. (2007). *HLM model building strategies: Class notes*. Retrieved from www.sarkisian.net/sc705/september6.pdf

Smith, A. F. M. (1973). A general Bayesian linear model. *Journal of the Royal Statistical Society, Series B, 54*, 61-75.
Snijders T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research, 22*, 342-363.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.
Tracz, S. M., Nelson, L., Newman, I., & Beltran, A. (2005). The misuse of ANCOVA: The academic and political implications of Type VI errors in studies of achievement and socioeconomic status. *Multiple Linear Regression Viewpoints, 31*, 16-21.

Send correspondence to: Susan Tracz
California State University, Fresno
Email: susant@csufresno.edu

APPENDIX A
SPSS Programming Instructions for HLM

To create variable for group centered means **SES means**
Data
Sort cases
Move s**chool** to box labeled "sort by"
OK
Data
Aggregate
Move **school** to box labeled "break variables"
Move **ses** to box labeled "summaries of variables"
C on box labeled "number of cases"
C on box labeled "add aggregate variables to active data base"
C on options for "file is already sorted on break variables"
OK

To create **SESdev**
Transform
Compute variable
Type **SESdev** in boy labeled "Target Variable"
Move variable labeled **SES** to box labeled "Numeric Expression"
Type * after **SES**
Move variable labeled **SES_mean** to box labeled "Numeric Expression"
C on OK

To run Model 1
Analyze
Mixed Models
Linear

Since **school** is the level 2 model in which individuals are nested, move **school** into the "Subjects" box
C on continue
Move **mathach** into the box labeled "Dependent variable"
C on the box on the left labeled "Fixed"
In the dropdown box in the center change "Factorial" to "Main Effects"
Continue
C on the box on the left labeled "Random"
- Move school from the box labeled "subjects" to "combinations"

- In the dropdown box in the center change “Factorial” to “Main Effects”
- C to check the box labeled “include intercept”
-In the dropdown box near the top change “Variance Components” to “Unstructured”

Continue

C on the box labeled “Statistics”

- C on “parameter estimates”
- C on “tests for covariance parameters”

Continue

OK

<u>To run Model 2</u>

Analyze

Mixed Models

Linear

Since **school** is the level 2 model in which individuals are nested, move **school** into the “Subjects” box

C on continue

Move **mathach** into the box labeled “Dependent variable”

Move **SESdev** into the box labeled “Covariates”

C on the box on the left labeled “Fixed”

- In the dropdown box in the center change “Factorial” to “Main Effects”
- Move **SESdev** from the box labeled “factors and covariates” to “Model”

Continue

C on the box on the left labeled “Random”

- Move school from the box labeled “subjects” to “combinations”
- In the dropdown box in the center change “Factorial” to “Main Effects”
- Move **SESdev** from the box labeled “factors and covariates” to “Model”
- C to check the box labeled “include intercept”
-In the dropdown box near the top change “Variance Components” to “Unstructured”

Continue

C on the box labeled “Statistics”

- C on “parameter estimates”
- C on “tests for covariance parameters”

Continue

OK

APPENDIX B

SPSS REML Multilevel Modeling Output for Models 1 and 2

Model 1: SPSS Output and Tabled Values

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	156.647	2673.663	.000

a. Dependent Variable: MATHACH.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	(γ_{00}) 12.636974	.244394	156.647	51.707	.000	12.154242	13.119706

a. Dependent Variable: MATHACH.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		(σ^2) 39.148322	.660645	59.258	.000	37.874662	40.464813
Intercept [subject = school]	Variance	(τ_{00}) 8.614025	1.078804	7.985	.000	6.739122	11.010548

Model 2: SPSS Output and Tabled Values

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	156.753	2670.923	.000
SESdev	1	155.217	292.403	.000

a. Dependent Variable: mathach.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	γ_{00} 12.636193	.244504	156.753	51.681	.000	12.153246	13.119140
SESdev	γ_{10} 2.193196	.128259	155.217	17.100	.000	1.939839	2.446554

a. Dependent Variable: mathach.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		(σ^2) 36.700200	.625744	58.650	.000	35.494030	37.947358
Intercept + SESdev [subject = school]	UN (1,1)	(τ_{00}) 8.680969	1.079535	8.041	.000	6.803240	11.076961
	UN (2,1)	(τ_{01}) .046792	.406372	.115	.908	-.749683	.843267
	UN (2,2)	(τ_{11}) .693989	.280786	2.472	.013	.314021	1.533721

a. Dependent Variable: mathach.

APPENDIX C

HLM Multilevel Modeling Output for Models 1 and 2

Model 1: HLM Output and Tabled Values

σ^2 = 39.14831

τ

INTRCPT1,β_0 8.61431

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.901

The value of the log-likelihood function at iteration 4 = -2.355840E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *df*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.636972	0.244412	51.704	159	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *df*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.636972	0.243628	51.870	159	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*df*	χ^2	*p*-value
INTRCPT1, u_0	2.93501	8.61431	159	1660.23259	<0.001
level-1, *r*	6.25686	39.14831			

Statistics for current covariance components model

Deviance = 47116.793469

Number of estimated parameters = 2

APPENDIX C (continued)

HLM Multilevel Modeling Output for Models 1 and 2

Model 2: HLM Output and Tabled Values

$\sigma^2 = 36.82835$

τ

INTRCPT1, β_0	4.82978	-0.15399
SES, β_1	-0.15399	0.41828

τ (as correlations)

INTRCPT1, β_0	1.000	-0.108
SES, β_1	-0.108	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, β_0	0.797
SES, β_1	0.179

The value of the log-likelihood function at iteration 21 = -2.331928E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *df*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.664935	0.189874	66.702	159	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.393878	0.118278	20.240	159	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *df*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.664935	0.189251	66.921	159	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.393878	0.117697	20.339	159	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*df*	χ^2	*p*-value
INTRCPT1, u_0	2.19768	4.82978	159	905.26472	<0.001
SES slope, u_1	0.64675	0.41828	159	216.21178	0.002
level-1, r	6.06864	36.82835			

Statistics for current covariance components model

Deviance = 46638.560919

Number of estimated parameters = 4