Empirical Characteristics of Centering Methods for Level-1 Predictor Variables in HLM

Randall E. Schumacker, Ph.D.
University of North Texas
Karen Bembry
Dallas Public Schools

Research has suggested that important research questions can be addressed with meaningful interpretations using hierarchical linear modeling. The proper interpretation of results, however, is invariably linked to the choice of centering for the Level-1 predictor variables which produce the outcome measures for the Level-2 regression analysis. In this study, three centering methods (uncentered, group mean, and grand mean) were compared using Read93 and Lunch Status as Level-1 predictor variables of ITBS94 reading test scores. The reliability estimates, or how accurately the sample estimate represents the population value, differed among the three centering methods. It was found that the group mean centering method provided the better reliability estimate. When using outcome measures based upon these three centering methods in a Level-2 analysis using two predictors, Gradrate and Percent Advdip, the group mean centering method indicated a more reliable estimate, but the grand mean centering method explained more between school variance. In fact, the gamma regression coefficients were markedly different, and the amount of variance explained was no longer consistent across the centering methods. These findings indicate that the choice of centering method for Level-1 predictor variables can affect empirical findings in HLM.

In quantitative research, it is essential that the variables under study are meaningful and interpretable so that statistical results can be related to theoretical concerns (Arnold, 1992). This principle is especially meaningful in multi-level analyses of variables such as in hierarchical linear modeling (HLM). In hierarchical linear modeling, the Level-1 variable s intercepts and slopes become outcome variables for Level-2 analyses. Because of potentially complex "nested" designs, it is important that each variables' value be clearly understood and specifically articulated (Bryk & Raudenbush, 1992).

Hierarchical linear modeling can be used to investigate many of the research questions in education that involve at least two levels of variables. Samples of such questions include: Do schools with a high percentage of students with limited English proficiency also have high achievement scores? Is the relationship between student SES and achievement invariant across schools? In fact, several studies investigating teacher effectiveness, school effectiveness, and student change and growth have been conducted using HLM (Bryk & Raudenbush, 1987 & 1988, Raudenbush, 1988, Lee & Bryk, 1989, Mendro et al. 1994, Webster et al, 1994). These studies recognize the nested design structure of students within classrooms, classrooms within schools, and schools within districts which produce different variance components for variables at each level.

In multi-level analyses, variables measured at the different levels provide different variance estimates (Bock, 1989), and depending on how the data are treated, opposing conclusions can be reached (Kreft, 1995; Kreft, de Leeuw, Aiken, 1995). For example, school level variables do not vary for students in a particular school. These school-level variables instead help to explain between-school variance rather than within-school variance. Likewise, students in the same classroom or school tend to be more alike than in other classrooms or schools; hence, the variance between students is not constant. Similarly, interpretations of outcomes can vary at the school-level, often leading to conflicting results. Student level data, however, measures the within-school variance, conditioned by school-level effects. In other words, the scores of students in each school building are adjusted using school-level variables, such as the crowded condition of that campus, to better reflect the nature and interpretation of the scores. A typical research question for an HLM analysis would be the investigation of the effect of a school s graduation rate and percent of students in advanced diploma plans on the mean reading test scores of 9th graders. In HLM terminology, this is a "means as outcomes" approach which involves an examination and use of the intercept values as outcomes (dependent variable) for Level-2 variable analysis. The ability to statistically analyze these characteristics within each school, until recently, has been overlooked. Most data analyses have been done using multiple regression single-level variable models.

One critical aspect to conducting HLM

analyses is centering Level-1 predictor variables that produce the outcomes that are used as dependent variables in Level-2 analyses. The interpretation of these outcomes is critical to the meaningfulness of results since centering changes, not only the coefficient s value, but also the research questions being answered by the statistical analysis (Burton, 1993). Theory should drive the decision to center any Level-1 variable as indicated by the research questions included in the investigation. This policy is in keeping with appropriate multiple linear regression procedures. With the introduction of HLM, however, the effect of one level of variables on another introduces several areas for further investigation (see conclusion section). The focus of this paper is on one such area, namely, centering effects of Level-1 variables.

Four possibilities exist for centering Level-1 predictor variables in HLM: X metric, grand mean, group mean, and user defined location, such as a cut-off score (Bryk & Raudenbush 1992). This study included the first three centering methods to determine whether the Level-1 centering decision affects the reliability estimates in the HLM analysis. This investigation further examined how centering decisions made for Level 1 variables affect the amount of between-school variance explained by Level-2 variables.

METHOD

Data Set

Research questions posed for this study were investigated using data from ninth grade students (n = 5638) continuously enrolled in 26 high schools within a large urban school district. The Level-1 variables in this study were defined as student level variables. The student level variables selected for this study included individual reading test scores from the Iowa Test of Basic Skills (ITBS94) for 1994 as the dependent variable. The 1993 individual reading scores (Read93) and an individual student socio-economic indicator identifying free-lunch status (Lunch Status) were the two independent predictor variables. The reading test scores were interval level data with a potential range from 0 to 26. Lunch Status was a dichotomous variable indicating whether or not a student was in the free lunch program.

Level-2 variables were defined as school-level variables. School level variables from the twenty-six high schools selected were the graduation rate for each high school (Gradrate) and the percent of the students in advanced diploma plans within each school (%Advdip). No Level-2 variables used in the study were aggregates of any individual Level-1 variables. Only the effects of the centering options on the "means as outcomes" or the intercept was investigated in this study.

Research Questions

Prior research has indicated that both an interpretation of intercept outcome values and a change in the research question occurs based upon a choice of centering method. Our concern, therefore, was not with theoretical issues which should be answered as an aspect of the research design, but with the empirical issues surrounding the reliability estimates. Thee reliability estimates represent how well the sample mean reflects the population mean and whether the amount of between-school variance predicted at Level-2 would be the same.

Analyses

Several analyses specifying different models were undertaken to answer the research questions. An initial analysis established a "fully unconditional" model, or a model without any Level 1 or Level 2 predictors (Bryk & Raudenbush, 1992). Two separate models with only a single Level 1 predictor variable were then specified. This was followed by a two predictor model with both variables included. A final analysis included a model with both Level-1 predictors (READ93, LUNCH) and two Level-2 predictors (AdvDip, Gradrate). Three analyses were run on each of these models. The analyses involved either an uncentered predictor, a predictor centered on the grand mean, or a predictor centered on the group mean. The Level-two predictors were not centered. The models are specified next.

Fully Unconditional Model

Student level (Level 1) $Y_{ij} = \beta_{oj} + r_{ij}$

where

Y_{ij} = ITBS 94 reading score for student I in school j

β_{oj} = mean reading score in school j

r_{ij} = Level-1 error, $N(0, \sigma^2)$; σ^2 = student level variance

School level (Level 2) $\beta_{oj} = \gamma_{oo} + u_{oj}$

where

β_{oj} = mean reading score in school j

γ_{oo} = grand mean of the district (N=26 schools)

u_{oj} = random effect school j, $N(0, \tau_{oo})$; τ_{oo} = school level variance

Level 1 Predictor Models

<u>READ93 model</u>

$Y_{ij} = \beta_{0j} + \beta_{1j}(READ93) + r_{ij}$

where
Y_{ij} = ITBS 94 reading score for student I in school j
β_{0j} = mean for school j
β_{1j} = slope for school j
r_{ij} = Level-1 error

$\beta_{0j} = \gamma_{00} + u_{0j}$

where
$_{00}$ = intercept mean of the district (n=26 schools)
u_{0j} = random effect for school j

$\beta_{1j} = \gamma_{10} + u_{1j}$

where
γ_{10} = slope mean of the district (n=26 schools)
u_{1j} = random effect for school j

Lunch model

$Y_{ij} = \beta_{0j} + \beta_{1j}(LUNCH) + r_{ij}$

where
Y_{ij} = ITBS 94 reading score for student I in school j
β_{0j} = mean for school j
β_{1j} = slope for school j
r_{ij} = Level-1 error

$\beta_{0j} = \gamma_{00} + u_{0j}$

where
γ_{00} = intercept mean of the district (n=26 schools)
u_{0j} = random effect for school j

$\beta_{1j} = \gamma_{10} + u_{1j}$

where
γ_{10} = slope mean of the district (n=26 schools)
u_{1j} = random effect for school j

Read93 and Lunch model

$Y_{ij} = \beta_{0j} + \beta_{1j}(READ93) + \beta_{2j}(LUNCH) + r_{ij}$

where
Y_{ij} = ITBS 94 reading score for student I in school j
β_{0j} = intercept for school j
β_{1j} = slope of READ93 for school j
β_{2j} = slope of LUNCH for school j
r_{ij} = Level-1 error

$\beta_{0j} = \gamma_{00} + u_{0j}$

where
γ_{00} = slope mean of the district (n=26 schools)
u_{0j} = random effect for school j

$\beta_{1j} = \gamma_{10} + u_{1j}$

where
γ_{10} = READ93 mean slope in the district
u_{1j} = random effect for school j

$\beta_{2j} = \gamma_{20} + u_{2j}$

where
γ_{20} = Lunch mean slope in the district
u_{2j} = random effect for school j

Level 1 and Level 2 Predictor Models

$Y_{ij} = \beta_{0j} + \beta_{1j}(READ93) + \beta_{2j}(LUNCH) + r_{ij}$

where
Y_{ij} = ITBS 94 reading score for student I in school j
β_{0j} = mean for school j
r_{ij} = Level-1 error

$\beta_{0j} = \gamma_{00} + \gamma_{01}(AdvDip) + \gamma_{02}(Gradrate) + u_{0j}$

where
γ_{00} = intercept mean of the district (n=26 schools)
u_{0j} = random effect for school j

$\beta_{1j} = \gamma_{10} + \gamma_{11}(AdvDip) + \gamma_{12}(Gradrate) + u_{1j}$

where
γ_{10} = Read93 slope mean of the district
u_{1j} = random effect for school j

$\beta_{2j} = \gamma_{20} + \gamma_{21}(AdvDip) + \gamma_{22}(Gradrate) + u_{2j}$

where
γ_{20} = Lunch slope mean of the district
u_{2j} = random effect for school j

The combined equation for the full model with both Level 1 and Level 2 predictor variables is then specified as:

$$Y_{ij} = (\gamma_{00} + \gamma_{01}(AdvDip) + \gamma_{02}(Gradrate) + u_{0j}) + (\gamma_{10} + \gamma_{11}(AdvDip) + \gamma_{12}(Gradrate) + u_{1j})(READ93) + (\gamma_{20} + \gamma_{21}(AdvDip) + \gamma_{22}(Gradrate) + u_{2j})(LUNCH) + r_{ij}$$

RESULTS

Level 1 Variable Analyses

The "fully unconditional" model, which only specified an intercept, resulted in a reliability estimate of .98 (Table 1). This initial "fully unconditional" null model allows us to partition the total variance in reading scores into a between school variance component (24%). It also establishes an estimate for the grand mean (β_0) , a confidence interval (β_0 +/- SEβ_0), and establishes the parameters for within-school variability (^{2}) and between school variability ($_{00}$). The reliability estimate indicates how well each school s sample average in reading achievement estimates their true mean (Bryk & Raudenbush, 1992). In this case, the reliability estimate was .98, indicating that the school s sample means are quite reliable as indicators of their true school means. The significant t-value indicates that the schools do not have the same mean ITBS 1994 reading average.

In the single Level 1 predictor model for READ93, results indicated that the reliability estimates differed between the three centering methods. The slope and reliability estimate, however, were the same as in the "fully unconditional" model. As expected, the amount of within school variance remained the same regardless of which centering method was used (45%).

Table 3 indicates results from the three centering methods when using Lunch as a single Level-1 predictor variable. The group mean centering method yielded results identical to the "fully unconditional" model, and the grand mean centering method more closely approximated this initial model than the uncentered approach. The amount of within school variance explained was small (3%), and as expected, the same regardless of choice of centering method.

Table 4 lists the results of the "fully unconditional" analysis and further indicates the effect of each centering method w n both Read93 and Lunch Status were used in a Level-1 prediction equation for 1994 ITBS reading outcomes. The results indicated that 45% of the between school variance was explained when using both predictors, which was the same amount indicated when using Re 93 alone, suggesting that the Lunch variable doesn't contribute any additional explained variance in the model. Moreover, the sample mean intercept value using the group mean centering method was the same as in the initial "fully unconditional" model, with only a slight improvement in the reliability estimate (.98 to .99). The reliability estimate for the grand mean centering method was more approximate to these values than the uncentered method, especially when using Read93 as the only predictor. The group mean centering method was therefore the most stable of the three centering methods.

From a practical research point of view, the choice of Level-1 predictors will impact the amount of within school variance explained. In our approach, preference would be given to using only Read93 as a Level-1 predictor since Lunch did not add any additional significant variance explained. However, for our purposes, we continued to use both Level-1 predictor variables in the Level-2 equation.

Level 1 and Level 2 variable analysis

Table 5 indicates each type of centering method and the associated summary statistics from the Level 2 complete model prediction equation. The amount of between-school variance explained is no longer consistent across the centering methods. The amount of variance explained using uncentered Level-1 variables was 89%; with group mean centering it was 70 %; and with grand mean centering it was 92%. The reliability estimates, or how well the sample estimates indicate the true population values, also differed. The group mean centering method yielded the highest reliability estimate (.96), but indicated very different coefficients for the variables than the other two centering methods, and had the lowest percent variance explained (70%). This leads to conflicting results since the group mean centering method was preferred in the Level 1 analyses, but the grand mean centering method explained more between-school variance in the Level 2 analysis.

Table 1. Full Unconditional Model on 1994 ITBS reading scores (n=26 schools).

Centering Method	β_0	SE β_0	r_{xx}
Null model	16.85	.67	.98

Note: No predictors were specified in Level 1 analysis; Intraclass correlation coefficient = $\tau_{00}/(\tau_{00} + {}^2) = 11.60/(11.60 + 36.08)$ = 24% of variance in 1994 ITBS reading scores explained between schools; and regression coefficient is significant (critical t =25.15, p=.0001).

Table 2. Level 1 predictor READ93 on 1994 ITBS reading scores (n=26 schools).

Centering Method[a]	β_0	SE β_0	r_{xx}	β_1	SE β_1	r_{xx}
Uncentered	13.85	.55	.69	1.87	.33	.67
Centered: Group Mean	16.85	.67	.98	1.82	.33	.67
Centered: Grand Mean	16.74	.62	.97	1.87	.33	.67

a
Intraclass correlation the same for each centering method [σ^2 (ANOVA) - σ^2 (READ93)/ σ^2 (ANOVA) = 36.08-20.00/36.08 = 45%

Table 3. Level 1 predictor Lunch on 1994 ITBS reading scores (n=26 schools).

Centering Method[a]	β_0	SEβ_0	r_{xx}	β_1	SE β_1	r_{xx}
Uncentered	13.85	.55	.69	1.87	.33	.67
Centered: Group Mean	16.85	.67	.98	1.82	.33	.67
Centered: Grand Mean	16.74	.62	.97	1.87	.33	.67

a
Intraclass correlation the same for each centering method [σ^2 (ANOVA) - σ^2 (Lunch)/ σ^2 (ANOVA) = 36.08-35.05/36.08 = 3%

Table 4. Both Level 1 predictor variables on 1994 ITBS reading scores (n=26 schools).

	Unconditional			Read93			Lunch		
Centering Method[a]	β_0	SE β_0	r_{xx}	β_1	SE β_1	r_{xx}	β_2	βSE $_2$	r_{xx}
Uncentered	6.46	.34	.47	.22	.004	.30	.69	.15	.25
Centered: Group Mean	16.85	.68	.99	.22	.004	.37	.69	.16	.28
Centered: Grand Mean	16.77	.26	.89	.22	.004	.29	.69	.15	.26

a
Intraclass correlation the same for each centering method [σ^2 (ANOVA) - σ^2(READ93 & Lunch)/ σ^2(ANOVA) = 36.08-19.88/36.08 = 45%

Table 5. Complete model: Graduation Rate and Percent Advanced Diploma using 1994 ITBS READ93 and Lunch reading score intercepts (n = 26 schools).

Centering Method	γ_{00}	SE γ_{00}	γ_{01}	SE γ_{01}	γ_{02}	SE γ_{02}	r_{xx}
Uncentered[a]	4.17	1.11	.003	.02	.04	.03	.43 3.75
Centered: Group Mean[b]	8.76	1.19	.030	.02	.13	.03	.96
Centered: Grand Mean[c]	14.1 5	.68	.004	.01	.05	.02	.84

[a] Intraclass correlation coefficient = τ_{00} (ANOVA) - τ_{00} (Gradrate & %Advdip)/ τ_{00}(ANOVA) = 11.60 - 1.22/ 11.60 = 89%; t = 3.75, p > .002.

[b] Intraclass correlation coefficient = τ_{00} (ANOVA) - τ_{00} (Gradrate & %Advdip)/ τ_{00}(ANOVA) = 11.60 - 3.46/ 11.60 = 70%; t = 7.36, p > .0001.

[c] Intraclass correlation coefficient = $_{00}$ (ANOVA) - $_{00}$ (Gradrate & %Advdip)/ τ_{00}(ANOVA) = 11.60 - .97/ 11.60 = 92%; t = 20.87, p > .00001.

CONCLUSIONS AND DISCUSSION

In practical applications, Level 1 predictor variables appear to become more stable when they are centered on either the group mean or grand mean. In our study, the initial sample estimate (intercept, β_0) was close to the population value in the "fully unconditional" model, as indicated by the reliability estimate of .98. This finding is expected in any initial null model. The reliability estimates, however, differed between the three centering methods when centering the Level 1 predictors Read93 and Lunch. For Read93, the reliability estimates were .76 uncentered), .98 (group-mean centered), and .90 (grand-mean centered). For Lunch, the reliability estimates were .69 (uncentered), .98 (group-mean centered), and .97 (grand-mean centered). The group-mean centering method for both Level 1 predictor variables yielded the same intercept and reliability estimate as in the "fully unconditional model". The intercept and slope values differed in the grand mean centering and uncentered methods, although they were more approximate when using grand mean centering. As expected, the amount of within school variance explained remained the same regardless of which centering method was used (45%). When using outcome measures based upon these three centering methods in a Level 2 full model analysis with two predictors, Gradrate and Percent Advdip, the group mean centering method also indicated a more reliable estimate, but the grand mean centering method explained more between school variance. The gamma regression coefficients were markedly different and the amount of variance explained was no longer consistent across the centering methods. These findings indicate that the centering of Level 1 variables empirically effects the variance estimation in Level 2 model analyses.

We found that the meaningfulness of the intercept and slope values in a Level 1 (student level) model depends upon the centering of the Level 1 predictor variables. In raw metric form, the equation $Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + r_{ij}$, yields intercept values, β_{0j}, which are interpreted as an outcome measure for a student attending school j who has a 0 (zero) on X_{ij}. Obviously this causes a problem in the interpretation of student achievement using these raw metric intercept values because the lowest score on the test will not be zero. When centering Level 1 predictor variables around the grand mean, they are determined by: $(X_{ij} - X_{..})$. The intercept, β_{0j}, can now be interpreted as an outcome measure for a student in school j whose value on X_{ij} is referenced to the grand mean. This permits a useful interpretation of the intercept as an adjusted mean for school j: in this case, $\beta_{0j} = \mu_{Yj} + \beta_{1j}(X_{.j} - X_{..})$. This is similar to the adjusted means in an ANCOVA analysis. These intercept values can now represent a specific interpretation of the outcome measures for each school in the Level 2 model analysis. The intercept variance term reflects the variation in the adjusted means for the set of schools. If the Level 1 predictor variables are centered around the group mean, they are determined by $(X_{ij} - X_{.j})$. Now the intercept, β_{0j}, represents the unadjusted outcome measure for a student in school j. In this instance,

$\beta_{0j} = \mu_{Yj}$. The intercept variance, Var (β_{0j}), is now the variance around the Level 2 variable unit means, μ_{Yj} . This permits an examination of the sampling distribution of the school means or slopes around a population mean value, i.e. district mean value.

A researcher will typically center some or all Level 1 (student-level) predictors at either the grand mean or group mean to add stability to the estimation process and provide for intercepts that can be meaningfully interpreted. Centering, however, also has the effect of changing the coefficients that are estimated and altering the research question(s) being asked. Burton (1993), using a NELS88 data set (outcome =mathematics achievement test; student-level variables =minority status, socio-economic status, and absenteeism; school-level variables =percent minority; location of school, and percent low SES), indicated that uncentered and grand mean centering indicated only significant Level 1 coefficients while group mean centering indicated significant Level 2 coefficients (school level). This implied two different interpretations of results: one at the student level with individual status affecting achievement, and one at the school level with average school status affecting achievement. It is troublesome that a choice between these two centering methods could result in two different interpretations. Which is the correct interpretation of the results?

Research has suggested that important research questions can be addressed with meaningful interpretations using hierarchical linear modeling (Raudenbush, Rowan, & Cheong, 1993). For practical applications, the unconditional model allows partitioning of variance into within-school and between-school components for the outcome measure. The choice of variables at Level-1 impacts the amount of within-school variance (student-level) that can be explained, and the choice of variables at Level-2 impacts the amount of between-school variance (school-level) that can be explained given the outcome measures provided from the Level-1 equation. The proper interpretation of results, however, is invariably linked to the choice of centering for the Level-1 predictor variables which produces the dependent measures for Level-2 regression analyses. Studies which examined organizational level, school effectiveness, and teacher effectiveness variables using hierarchical linear models have provided more appropriate variance estimates and means as outcomes than previous single level data analyses. The proper interpretation and accuracy of estimation, however, requires that a researcher pay special attention to the centering effects in Level-1 student-level variables upon Level-2 analyses when conducting hierarchical linear models.

For many researchers, multiple regression has become a valuable data analytic tool because many of the issues related to using multiple regression have been investigated. For example, sample size and power, non-normality, heterogeneity, number of predictors, ratio of sample size to predictors, multi-collinearity, use of composite variables, outliers, and interaction effects. We believe that many of these concerns need to be restated in the context of hierarchical linear modeling. Once case in point is the effect of centering when including an interaction term. Aiken & West (1993) have indicated that centering variables in the presence of an interaction term in multiple regression changes the value of the regression coefficients. In HLM, this would follow as a dictum, especially in light of the findings by Burton (1993). Additional examination of other factors will determine what effect, if any, they have upon hierarchical linear analyses.

REFERENCES

Aiken, L. S. & West, S. G. (1991). Multiple *Regression: Testing and Interpreting Interactions.* Newbury Park, CA: Sage Publications.

Arnold, C. L. (1992). An introduction to hierarchical linear models. *Measurement and Evaluation in Counseling and Development, 25,* 58-90.

Bock, R. Darrell. (1989). *Multilevel Analysis of Educational Data.* San Diego, CA: Academic Press.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications, and Data Analysis Methods.* Newbury Park, CA: Sage Publications.

Bryk, A. S. & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education,* November, 65-108.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147-158.

Burton, Bob. (1993, April). Some observations on the effect of centering on the results obtained from hierarchical linear modeling. (Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA).

Kreft, I.G.G. (1995, April). The effects of centering in multilevel analysis: Is the Public school the loser or the winner? A new analysis of an old question. Paper presented at the American Educational Research Association annual meeting. San Francisco, California.

Kreft, I.G.G., de Leeuw, J., Aiken, L.S. (1995). The effect of different forms of centering in hierarchical linear modeling. Multivariate Behavioral Research, in press.

ee V. E. & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 62, 172-192.

Mendro, R. L., Webster, W. J., Bembry, K. L. & Orsak, T. H. (1994, October). Applications of Hierarchical Linear Models in Identifying Effective Schools. (Paper presented at the Annual Meeting of the Rocky Mountain Educational Research Association, Tempe, AZ).

Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13, 85-116.

Webster, W. J., Mendro, R. L., Bembry, K. L. & Orsak, T. H. (1994, October). Alternative methodologies for identifying effective schools (Paper presented at the Annual Meeting of the Rocky Mountain Educational Research Association, October, Tempe, AZ).