Psychological Methods
2007, Vol. 12, No. 2, 121–138

1082-989X/07/$12.00 DOI: 10.1037/1082-989X.12.2.121

Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue

Craig K. Enders and Davood Tofighi
Arizona State University

Appropriately centering Level 1 predictors is vital to the interpretation of intercept and slope parameters in multilevel models (MLMs). The issue of centering has been discussed in the literature, but it is still widely misunderstood. The purpose of this article is to provide a detailed overview of grand mean centering and group mean centering in the context of 2-level MLMs. The authors begin with a basic overview of centering and explore the differences between grand and group mean centering in the context of some prototypical research questions. Empirical analyses of artificial data sets are used to illustrate key points throughout. The article provides a number of practical recommendations designed to facilitate centering decisions in MLM applications.

Keywords: centering, grand mean centering, group mean centering, multilevel models, hierarchical linear models

Supplemental materials: http://dx.doi.org/10.1037/1082-989X.12.2.121.supp

Multilevel models (MLMs) have enjoyed widespread use in the behavioral sciences in recent years. These models are attractive because they provide a straightforward mechanism for analyzing data sets in which study participants are clustered within organizational units (e.g., clients nested within therapists, students nested within schools, individuals nested within families, employees nested within companies, daily diary observations nested within individuals). In contexts such as these, it is often of interest to explore the influence of variables at both levels of the data hierarchy, and MLMs readily allow one to examine the influence of individual (i.e., Level 1) and cluster-level (i.e., Level 2) covariates.

Psychological constructs are frequently expressed on arbitrary metrics that lack a clearly interpretable or meaningful zero point (Blanton & Jaccard, 2006). Although it is clearly not a panacea for the problem of arbitrary metrics, centering can be used to establish a zero point on scales that otherwise lack such a value. The use of centering for this purpose is relatively straightforward in ordinary least squares (OLS) regression (Aiken & West, 1991), but it is considerably more complex when considering Level 1 predictors in an MLM analysis. In the MLM context, Level 1 covariates can be centered at the grand mean (centering at the grand mean; CGM), or they can be deviated around the mean of the cluster *j* to which case *i* belongs; the latter option is frequently referred to as *group-mean centering* in the MLM literature, but we henceforth refer to this technique as *centering within cluster* (CWC) in order to avoid confusion with CGM. Although both forms of centering can be used to establish a meaningful zero point, CGM and CWC generally produce parameter estimates that differ in value and also in meaning. Although there are unique situations in which CGM and CWC produce equivalent parameter estimates (Kreft, de Leeuw, & Aiken, 1995), this is the exception rather than the norm.

It should be noted that the centering of Level 2 (e.g., organizational level) variables is far less complex than the centering decisions required at Level 1, as it is only necessary to choose between the raw metric and CGM; CWC is not an option because each member of a given cluster shares the same value on the Level 2 predictor. Centering decisions at Level 2 generally mimic prescribed practice from the OLS regression literature (Aiken & West, 1991), so the focus of this article is on centering at Level 1. Throughout the remainder of the article, we assume that all Level 2 predictors are centered at their grand mean.

The issue of centering predictor variables in MLM has been discussed in the methodological literature (Hofmann &

Craig K. Enders and Davood Tofighi, Department of Psychology, Arizona State University.

Correspondence concerning this article should be addressed to Craig K. Enders, Department of Psychology, Box 87-1104, Arizona State University, Tempe, AZ 85287-1104. E-mail: cenders@asu.edu

Gavin, 1998; Kreft, 1995; Kreft et al., 1995; Longford, 1989b; Paccagnella, 2006; Plewis, 1989; Raudenbush, 1989a, 1989b; Wu & Wooldridge, 2005), but it is still widely misunderstood by substantive researchers and methodologists alike. We conducted an informal review of recent volumes of several American Psychological Association (APA) journals and found examples of inappropriate centering practices. For example, there were cases in which the authors made no mention of the centering technique they used (Griffin, 2001; Kallestad & Olweus, 2003), making the resulting regression coefficients difficult, if not impossible, to interpret. In other cases, authors explicitly stated that the Level 1 predictors were left uncentered (Balsam, Beauchaine, Mickey, & Rothblum, 2005), even though a number of these variables had a raw metric with no meaningful zero point (e.g., age, education, self-esteem, life satisfaction). It is interesting to note that we also found cases in which authors based centering decisions on the results of empirical tests rather than on their substantive research questions. For example, Wampold and Brown (2005) stated that "We conducted the analysis using both parameterizations [CGM and CWC], which yielded the finding that the within-therapist regression coefficient was not too different from the between-therapist regression coefficient" (p. 918). Similarly, Zohar and Luria (2005) explained that "When testing moderation effects in this and subsequent hypotheses, we first tested the effect of centering of variables in the statistical models . . . This operation resulted in no improvement, and thus it was not performed in the final model" (p. 623). The decision to base centering decisions on empirical results is problematic because it should be based on the substantive question of interest. Kreft et al. (1995) underscored this point, stating that "There is no statistically correct choice among RAS [the raw metric], CGM, and CWC" (p. 17). It is certainly not our intent to single out these articles as exemplars of poor practice, nor is it our intent to criticize these authors. Rather, we use these articles to illustrate the complexity of centering as a methodological issue and to demonstrate that confusion exists, even in articles published in top APA journals.

Given the confusion that exists over centering, the purpose of this article is to provide a detailed overview of CGM and CWC and to provide a number of practical recommendations designed to facilitate centering decisions in MLM research. In doing so, we chose to avoid comparisons of CGM and CWC based on numerical stability or computational efficiency (e.g., Longford, 1989a). Instead, we attempt to address a relatively simple question: Which form of centering provides interpretable parameter estimates that can be used to address a particular substantive research question?

Our discussion of centering is restricted to the cross-sectional case, and we do not consider the issue of centering in longitudinal MLMs (i.e., growth curve models). In the typical growth curve application, the Level 1 covariate of interest is a temporal predictor such as age or elapsed time. In situations such as this, the Level 1 predictor (i.e., the "time" variable) is typically centered around a fixed value rather than around the mean. For example, in a longitudinal study of adolescent depression, a child's age at a particular assessment (i.e., the Level 1 predictor) could be expressed as a deviation relative to age 16 (e.g., Montague, Enders, Dietz, Morrison, & Dixon, 2006), and this centered age variable would appear in the Level 1 regression model. Centering around a fixed point in time in a longitudinal study is arguably less complex than the decision to use CGM or CWC in the cross-sectional context, so we focus our discussion on the latter situation. Excellent discussions of centering in the context of growth curve models are found in Singer and Willett (2003) and in Biesanz, Deeb-Sossa, Aubrecht, Bollen, and Curran (2004). However, our discussion of centering does apply to other types of within-person data structures in which multiple observations are obtained from each individual but in which intraindividual variation is not expressed as a function of elapsed time (e.g., daily diary observations nested within individuals). Interested readers can consult Nezlek (2001) for a discussion of multilevel modeling in this context.

Illustrative Data

In order to make our discussion more concrete, we rely on a hypothetical research scenario from the occupational stress literature. Specifically, suppose it was of interest to study the influence of workload (measured by the number of work hours per week; *HOURS*) on psychological well-being (*WELLBEING*). Furthermore, assume that the data are hierarchically structured such that employees are nested within workgroups (organizations). In this scenario, both workload and well-being are Level 1 (i.e., individual level) variables, but it may also be of interest to investigate the influence of organizational-level (i.e., Level 2) variables such as workgroup size (*SIZE*). Our decision to use an occupational stress example was somewhat arbitrary, and the concepts discussed in this article generalize to any number of multilevel analysis problems. Nevertheless, we felt that this research scenario was useful because it provides an intuitive platform from which to explore centering issues.

The Level 1 regression equation for the illustrative model is

$$WELLBEING_{ij} = \beta_{0j} + \beta_{1j}(HOURS_{ij}) + r_{ij}, \tag{1}$$

where β_{0j} is the intercept for cluster j, β_{1j} is the regression coefficient for cluster j, and r_{ij} is the Level 1 residual.[1] In turn, the Level 2 models for the intercept and slope are as follows:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}. \tag{3}$$

The equations above express each cluster's intercept and slope as a function of the mean intercept and slope (γ_{00} and γ_{10}, respectively) plus a residual term that captures cluster j's deviation from the mean (u_{0j} and u_{1j}).

In principle, each cluster has a unique slope and intercept, but the MLM analysis summarizes the j regressions by using a smaller set of parameters. Substituting Equations 2 and 3 into the Level 1 equation yields the so-called combined model regression equation shown below.

$$WELLBEING_{ij} = \gamma_{00} + \gamma_{10}(HOURS_{ij}) + u_{0j} + u_{1j}(HOURS_{ij}) + r_{ij}. \tag{4}$$

This basic model yields six parameter estimates: the mean intercept and slope (γ_{00} and γ_{10}, respectively), variance estimates for u_{0j} and u_{1j} that quantify the heterogeneity in the intercepts and slopes (τ_{00} and τ_{11}, respectively), the covariance between the intercepts and slopes (τ_{10}), and the Level 1 residual variance (σ^2).

To facilitate the subsequent discussion, we generated a small artificial data set that consisted of three clusters (workgroups) and $n_j = 5$ cases (employees) within each cluster. A scatter plot displaying the raw data is given in the top panel of Figure 1. The data were generated such that work hours and well-being contained a considerable proportion of between-cluster variation (i.e., mean differences among clusters accounted for a significant portion of the total score variation). The between-cluster variation in work hours is evidenced by the spread of the clusters along the horizontal axis, and the vertical separation of the clusters reflects between-group variation in the well-being scores. The mean differences in workload accounted for approximately 69% of the total score variation (i.e., the intraclass correlation, or the proportion of the total variation that exists at Level, 2 was $\tau_{00}/[\tau_{00} + \sigma^2] = .69$), and differences among well-being means accounted for 84% of the variation. These data do not represent a realistic application of MLM because the number of clusters is very small, and the intraclass correlation is larger than what one might typically see with cross-sectional data. However, these characteristics make the artificial data set useful for visually demonstrating the effects of different forms of centering.

It may not be immediately obvious, but the total correlation between workload and well-being has both a within- (i.e., Level 1) and a between-cluster (i.e., Level 2) component.[2] Returning to the top panel in Figure 1, the scores within each workgroup are negatively correlated, such that individuals who work more tend to have lower well-being scores. Additionally, the workload and well-being means for each cluster can be viewed as Level 2 variables (Paccagnella, 2006), the scatter plot for which is shown in the bottom panel of Figure 1. As seen in the figure, the cluster means are nearly perfectly correlated, such that the average well-being score for a particular workgroup decreases as the mean number of work hours increases. Figure 1 depicts an association between workload and well-being at both levels of the hierarchy, yet this complex relationship is summarized with a single regression slope in Equation 4. This is the crux of the centering issue in MLMs, and the goal of this article is to demonstrate how CGM and CWC partition the relationship between X and Y and produce different interpretations of the MLM parameters.

Centering at the Grand Mean (CGM)

Under CGM, the Level 1 predictor (work hours) is deviated around the grand mean (i.e., $HOURS_{ij} - \bar{x}_{HOURS}$); for simplicity, the grand mean-centered scores are henceforth referred to as $HOURS_{cgm}$. Before considering the impact of centering on MLM parameters, it is informative to see how CGM affects the multilevel structure of the data. Applying CGM to the small artificial data set produced the scatter plot shown in the top panel of Figure 2. As seen in the figure, the CGM scatter plot is identical to that of the raw data (see Figure 1), except for the labeling of the workload values along the horizontal axis. The fact that CGM does nothing to the rank order of scores on either variable suggests that the complex, multilevel association between work hours and well-being is unaffected by grand mean centering.

To further illustrate the impact of CGM, the correlations and cluster means from the artificial data are given in Table 1. As seen in the table, the magnitude of the mean differences on the workload variable is unaffected by CGM, and the raw and centered workload scores possess the same correlation with the well-being ($r = -.86$). More important,

[1] A brief explanation of the notation used in this article is warranted. Like OLS regression, the values of explanatory variables in MLMs are assumed to be fixed as opposed to random. Our notational scheme makes no distinction between fixed and random variables, and we generically refer to any measured variable by using uppercase, italic typeface. The realized values of a variable are designated using lowercase, italic typeface (e.g., $x_{ij} = 0$). Finally, all residual terms are denoted using lowercase, italic typeface.

[2] Interested readers are encouraged to consult Snijders and Bosker (1999, p. 26) for an excellent discussion of within- and between-cluster relationships.

the table shows that $HOURS_{cgm}$ is correlated with both Level 1 and Level 2 variables (e.g., workgroup size; *SIZE*). At first glance, it may appear strange that a Level 1 variable is correlated with variables at both levels of the hierarchy. However, the fact that the intraclass correlation is greater

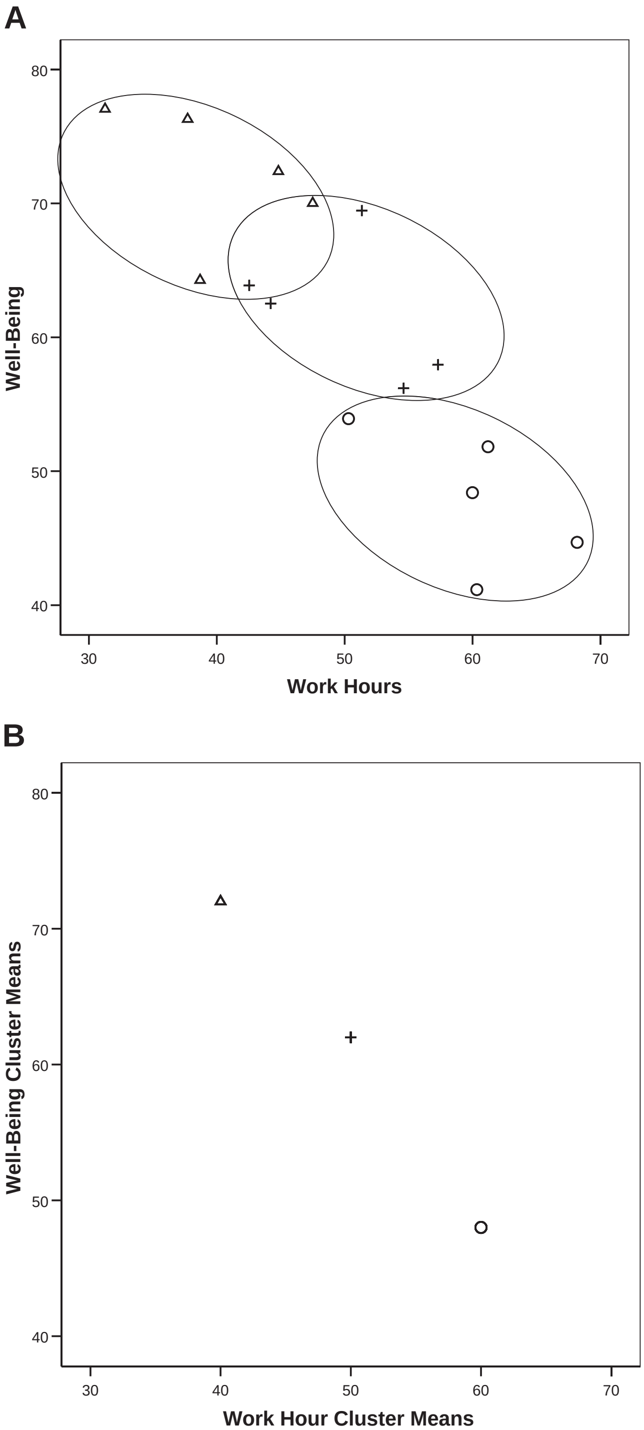

Figure 1. A: Scatter plot of the artificial data set. Cluster 1 cases are denoted by open triangles, and Cluster 2 and 3 cases are denoted by plus signs and open circles, respectively. A negative correlation exists within each cluster, such that higher workload scores are associated with lower well-being. B: Scatter plot of the cluster means. A negative association exists among the means, such that the average well-being score decreases as the mean workload increases.

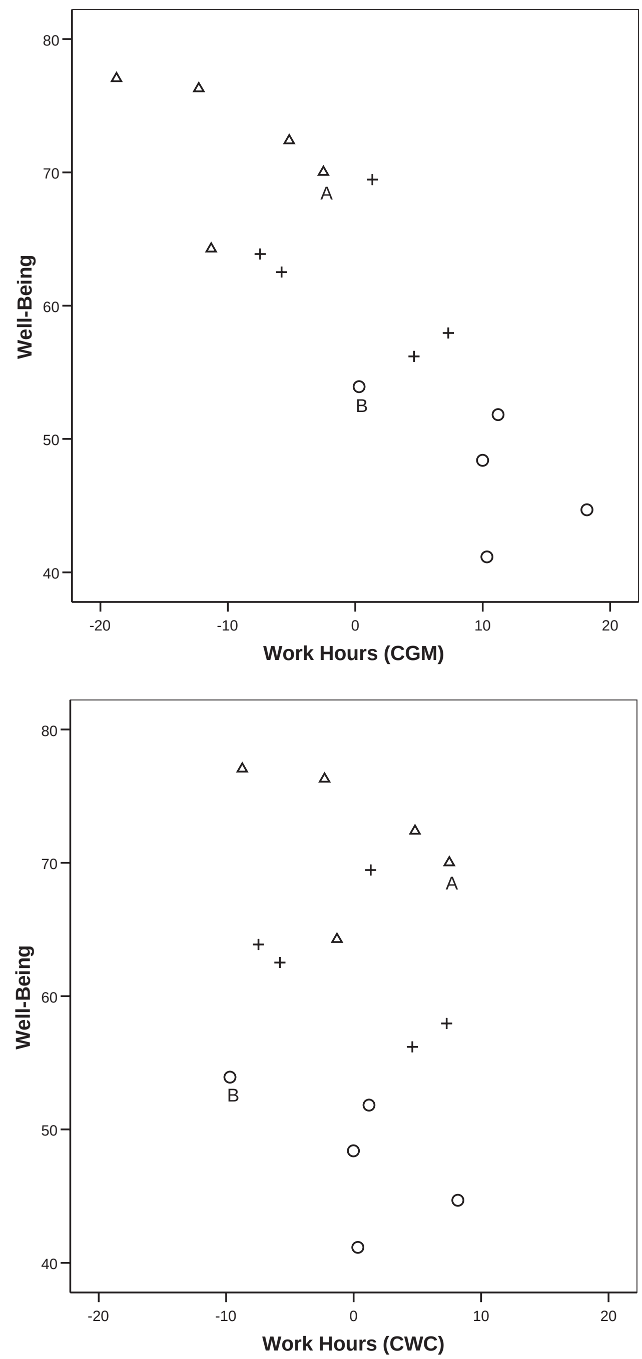

Figure 2. Scatter plot of artificial data under grand mean centering (CGM) and within-cluster centering (CWC). Cluster 1 cases are denoted by open triangles, and Cluster 2 and 3 cases are denoted by plus signs and open circles, respectively.

Table 1
Cluster Means and Correlations for the Artificial Data Under Different Forms of Centering

Cluster	Cluster *M*			
	WELLBEING	$HOURS_{raw}$	$HOURS_{cgm}$	$HOURS_{cwc}$
1	72	40	−10	0
2	62	50	0	0
3	48	60	10	0

Data	1	2	3	4	5	6	7
1. *WELLBEING*	—						
2. $HOURS_{raw}$	−.86	—					
3. $HOURS_{cgm}$	−.86	1.00	—				
4. $HOURS_{cwc}$	−.22	.57	.57	—			
5. $\bar{x}_{HOURS_j}$	−.90	.82	.82	0	—		
6. $\bar{y}_{WELLBEING_j}$	.90	−.82	−.82	0	−.99	—	
7. *SIZE*	−.44	.33	.33	0	.40	−.48	—

Note. RAW = original metric; CGM = grand mean centered; CWC = centered within cluster; WELLBEING = psychological well-being; *HOURS* = workload measured in hours per week; *SIZE* = workgroup size.

than zero (and less than unity) implies that work hours (and thus $HOURS_{cgm}$) can be viewed as a composite variable that contains both within- and between-cluster variation (i.e., $HOURS_{ij} - \bar{x}_{HOURS} = [HOURS_{ij} - \bar{x}_{HOURS_j}] + [\bar{x}_{HOURS_j} - \bar{x}_{HOURS}]$ = within + between). The presence of variation at both levels of the hierarchy leads to an important conclusion: CGM yields scores that are correlated with variables at both levels of the hierarchy. The preceding point is vital for understanding the differences between CGM and CWC, and we revisit it throughout the article.

How does CGM affect the interpretation of the MLM parameters? The Level 1 regression equation under CGM is

$$WELLBEING_{ij} = \beta_{0j} + \beta_{1j}(HOURS_{ij} - \bar{x}_{HOURS}) + r_{ij}. \quad (5)$$

Equation 5 shows that the intercept for cluster *j* is the predicted score for a case whose workload value is at the grand mean (i.e., when $HOURS_{ij} = \bar{x}_{HOURS}$, the predicted *WELLBEING* score is β_{0j}). Further insight into the definition of β_{0j} can be achieved by taking the expectation of Equation 5 within each cluster (the expectation of $WELLBEING_{ij}$ and $HOURS_{ij}$ is the cluster mean, and the expectation of r_{ij} is zero), and rearranging terms as follows:

$$\beta_{0j} = \mu_{WELLBEING_j} - \beta_{1j}(\bar{x}_{HOURS_j} - \bar{x}_{HOURS}). \quad (6)$$

Equation 6 shows that the Level 1 intercept is equal to the well-being mean for cluster *j*, minus an adjustment that depends on the regression slope, and on the deviation between the workload mean for cluster *j* and the grand mean. Readers may recognize Equation 6 as the formula used to obtain adjusted means in the analysis of covariance (ANCOVA) context, and the Level 1 intercept can also be interpreted as the adjusted mean for cluster *j* (i.e., the expected well-being score for a case belonging to cluster *j* after "equating" the clusters on their average workload). A graphical depiction of the ANCOVA adjustment can be found in intermediate statistics texts (e.g., Stevens, 1999, p. 311).

As discussed earlier, the MLM analysis ultimately yields an estimate of the mean intercept, which was denoted as γ_{00} in Equation 4. From the previous discussion, it follows that this parameter is interpreted as the average adjusted mean. Although the MLM intercept has a straightforward interpretation, the presence of between-cluster variation in the centered workload scores makes the interpretation of the CGM slope coefficient more problematic. To illustrate, the top panel of Figure 3 superimposes three regression lines over the CGM scatter plot. Specifically, the dotted line represents the regression of well-being on workload within each cluster (i.e., the pooled within-cluster regression, β_w). Next, the solid line reflects the relationship between the cluster means at Level 2 (i.e., the between-cluster regression, β_b); this association was previously depicted in the bottom panel of Figure 1. Finally, the dashed line represents the total regression line (β_t) and can be thought of as the slope obtained from an analysis that ignores the nesting of cases within organizations (i.e., a disaggregated analysis).

When analyzing clustered data such as these using OLS regression, the total regression coefficient is actually a weighted combination of the within- and between-cluster regression coefficients (Raudenbush & Bryk, 2002, p. 137). This result is clearly evident in the top panel of Figure 3, in which the total regression line (the dashed line) is flatter than the between-cluster slope (the solid line) but is steeper than the within-cluster regression (the dotted line). The problem with an OLS analysis of clustered data is that it incorrectly summarizes the complex association between *X* and *Y* by using only a single parameter (β_t). As a result, "$\hat{\beta}_t$ is generally an uninterpretable blend of $\hat{\beta}_w$ and $\hat{\beta}_b$" (Raudenbush & Bryk, 2002, p. 138).

Returning to the combined model notation in Equation 4, the association between workload and well-being is also represented by a single regression coefficient (γ_{10}) that is analogous to β_t. The presence of between-cluster variation in the centered scores means that the CGM regression slope is also a mixture of the within- and between-cluster association between workload and well-being. Raudenbush and Bryk (2002) made this point, stating that, "the hierarchical estimator under grand-mean centering is an inappropriate estimator of the person-level [i.e., Level 1] effect. It too is an uninterpretable blend: neither β_w nor β_b" (p. 139).

The use of CGM also affects the interpretation of the MLM variance components. We have previously shown that the intercept for cluster *j* is interpreted as an adjusted mean, so it follows that the CGM estimate of the intercept variance (τ_{00}) quantifies variation in the adjusted outcome means (i.e., the variation in the well-being means after

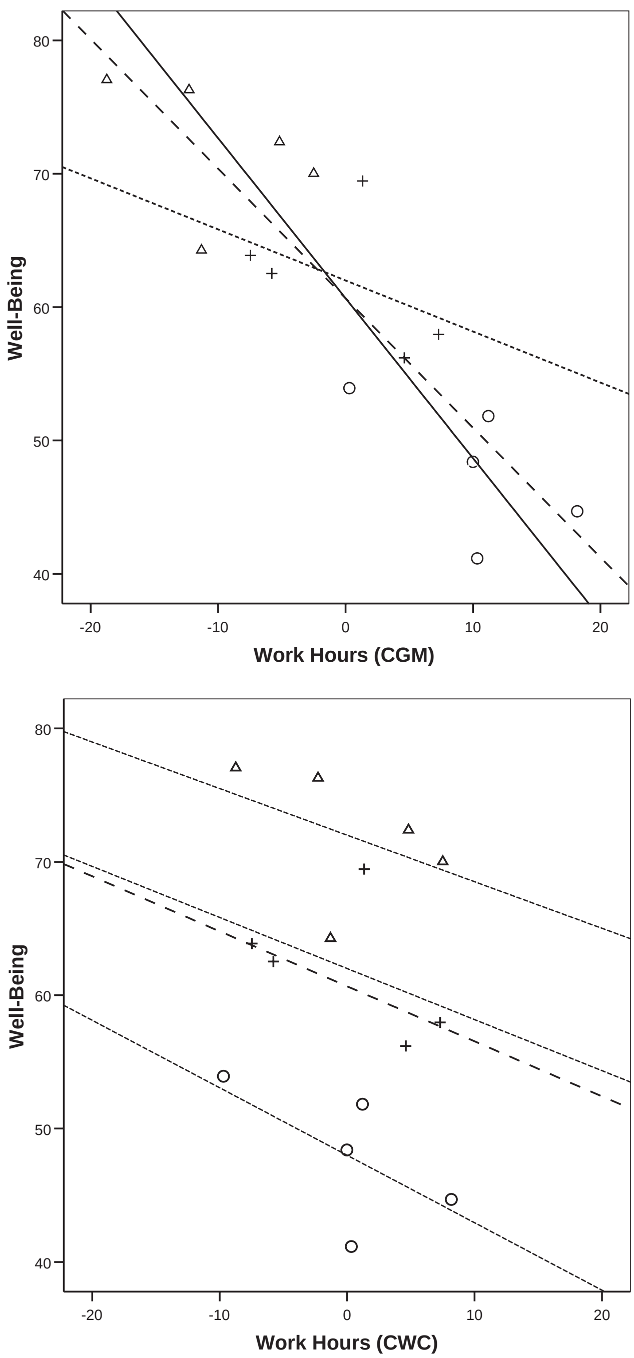

Figure 3. Scatter plot of artificial data under grand mean centering (CGM) and within-cluster centering (CWC). Ordinary least squares regression lines are superimposed over the plot. In the top panel (CGM), the dotted line represents the pooled within-cluster regression (β_w), whereas the solid line is the between-cluster regression of the well-being means on the workload means (β_b). The dashed line is the total regression that ignores the clustering of cases within organizations (β_t). In the bottom panel (CWC), the workload scores contain no between-cluster variation, so $\beta_b = 0$, and the total regression line quantifies the pooled within-cluster regression coefficient.

partialling out the effect of workload). However, the CGM estimate of the slope variance (τ_{11}) is problematic and is potentially biased toward zero. When the regression slopes vary across clusters, a dependency is introduced between the intercepts and slopes, because the magnitude of the adjustment to each cluster's mean (see Equation 6) depends on the cluster-specific slope coefficient (β_{1j})—this is analogous to violating the homogeneity of regression assumption in ANCOVA. This dependency serves to confound the CGM estimates of the intercept and slope variance, making the interpretation of these estimates somewhat ambiguous.

Understanding the resulting bias in the CGM slope variance estimate requires a brief tangent into the expectation maximization (EM) algorithm, an iterative computational algorithm used to obtain maximum likelihood parameter estimates. The EM algorithm requires the computation of empirical Bayes (EB) estimates of each cluster's intercept and slope, computed as a weighted sum of the OLS estimates of β_{0j} and β_{1j} and the model-predicted values of the intercept and slope (e.g., γ_{00} and γ_{10}). When the range of X values for cluster j does not include the grand mean, the OLS estimate of the intercept is unreliable because the adjustment term in Equation 6 involves extrapolation outside the range of observed X values. As a consequence, the EB estimate of the intercept is more heavily weighted by a constant model-predicted value, and the estimate of the intercept variation becomes attenuated—this concept is referred to as *shrinkage* in the MLM literature. Because CGM induces a dependency between the intercepts and slopes, it follows that the variation in the slopes is also artificially compressed, resulting in a negatively biased estimate of the slope variance (τ_{11}).

Centering Within Cluster (CWC)

Under CWC, the Level 1 predictor (work hours) is deviated around the mean of the cluster j to which case i belongs (i.e., $HOURS_{ij} - \bar{x}_{HOURS_j}$); for simplicity, the group mean-centered scores are henceforth referred to as $HOURS_{cwc}$. Again, it is informative to see how CWC affects the multilevel structure of the data before considering its impact on MLM parameters. The bottom panel of Figure 2 shows a CWC scatter plot that is radically different from that of CGM. To illustrate, consider the data points labeled A and B in the figure. Case A has the highest workload among the Cluster 1 cases (denoted with a diamond), and Case B has the lowest workload in Cluster 3 (denoted with a circle). Furthermore, the top panel of Figure 2 shows that Case B has a slightly higher workload score than Case A (this is also true of the raw data, as seen in Figure 1). In contrast, the relative position of these two cases changes rather dramatically under CWC, such that Case A has the higher workload score. The rank order change owes to the fact that workload scores are now expressed relative to other cases belonging to the same cluster.

The descriptive statistics and correlations in Table 1 further clarify the differences between CGM and CWC. Unlike CGM, CWC fundamentally alters the mean and correlation structure of the data. For example, note that the workload cluster means are all zero under CWC; said differently, $HOURS_{cwc}$ contains no between-cluster variation. The absence of between-cluster variation is seen graphically in the bottom panel of Figure 2, in which the clusters are no longer separated along the horizontal axis.

Removing the between-cluster variation in the workload scores has important implications for the correlation structure. Specifically, notice that $HOURS_{cwc}$ is correlated with Level 1 variables (e.g., well-being) but is uncorrelated with Level 2 variables (e.g., workgroup size). Also, note that the magnitude of the correlation between workload and well-being is different under CGM and CWC. $HOURS_{cgm}$ is a composite variable that contains variation at both levels of the hierarchy, so it follows that the correlation between $HOURS_{cgm}$ and well-being is a blend of the within- and between-cluster associations depicted in Figure 1. In contrast, the correlation between $HOURS_{cwc}$ and well-being is much weaker ($r = -.22$ versus $-.86$) because it is unaffected by the strong Level 2 association shown in the bottom panel of Figure 1. The fact that CWC purges scores of all between-cluster mean differences leads to a second important conclusion: CWC produces scores that are uncorrelated with Level 2 variables. Again, the preceding point is vital for understanding the differences between CGM and CWC and is revisited throughout the remainder of the article.

Having gained some intuition about CWC, we now examine its impact on the interpretation of MLM parameters. The Level 1 regression equation under CWC is

$$WELLBEING_{ij} = \beta_{0j} + \beta_{1j}(HOURS_{ij} - \bar{x}_{HOURS_j}) + r_{ij}. \quad (7)$$

Equation 7 shows that the intercept is the predicted score for a case in which the workload value is at the cluster mean (i.e., when $HOURS_{ij} = \bar{x}_{HOURS_j}$, the predicted *WELLBEING* score is β_{0j}). From OLS regression, we know that the mean of Y is the predicted value associated with the mean of X, so it follows that the Level 1 intercept is the unadjusted mean for cluster j. By extension, the intercept parameter in the combined model (γ_{00} in Equation 4) is interpreted as the average unadjusted cluster mean.

The use of CWC also changes the interpretation of the MLM regression slope. The CGM slope coefficient is problematic because it is a mixture of the within- and between-cluster association between X and Y. In contrast, CWC removes all between-cluster variation from the predictor variable and yields a slope coefficient (i.e., the γ_{10} coefficient in Equation 4) that is unambiguously interpreted as the pooled within-cluster (i.e., Level 1) regression of well-being on workload. This can be seen graphically in the bottom panel of Figure 3, in which the total regression line (the dashed line) is approximately equal to the pooled within-cluster regressions (the dotted lines).

The use of CWC also affects the interpretation of the variance components. Because the Level 1 intercept is an unadjusted mean, it follows that the intercept variance (τ_{00}) quantifies the between-cluster variation in the outcome scores, or the variance of the unadjusted cluster means. As such, the estimate of τ_{00} obtained from CWC should be nearly identical to that obtained from an unconditional model with no predictors. We previously pointed out that the CGM estimate of the slope variance is potentially biased because of a dependency on intercepts that were subject to shrinkage. The same is not true of CWC, so the resulting estimate of the slope variation is generally more accurate (Raudenbush & Bryk, 2002).

The Linkage Between Centering and Substantive Questions

Thus far, we have established that the interpretation of MLM parameters is fundamentally different under CGM and CWC. To briefly summarize, CGM produced an intercept value that was interpreted as the average adjusted mean, whereas the CWC intercept value was the average unadjusted mean. The centered scores under CGM contained both within- and between-cluster variation, resulting in a regression slope that was an ambiguous mixture of the Level 1 (e.g., person level) and Level 2 (e.g., organization level) association between X and Y. In contrast, centered scores under CWC were uncorrelated with Level 2 variables, so the resulting regression coefficient was a pure estimate of the Level 1 relationship between X and Y. The choice of centering also affected the interpretation of the variance components. The CGM estimate of the intercept variance quantified variation in the adjusted means (i.e., variation in the outcome means, having controlled for the Level 1 predictor), whereas the CWC variance estimate captured variation in the unadjusted means. We also discussed the differences in the estimates of the slope variation and pointed out potential problems associated with the CGM estimate of the slope variance.

From the previous discussion, one might reasonably conclude that the use of CGM is problematic and should be avoided. However, this is far from true. It is our view that the choice of centering method is intimately linked to one's substantive research questions, and both CGM and CWC are appropriate in certain circumstances and are inappropriate in others. It is generally true that a particular research question calls for either CGM or CWC, but it is also true that CGM and CWC may be used to address different questions within the same study.

Given the linkage between centering and one's substantive question, we further explore the differences between CGM and CWC in the context of some prototypical research questions that may be of interest to researchers who use MLM. In doing so, we hope to provide a few straightforward rules of thumb for centering Level 1 predictors. For this purpose, we classify substantive research questions into four broad scenarios: (a) the primary substantive interest involves a Level 1 predictor (e.g., What is the impact of an individual's workload on his or her well-being?), (b) the primary substantive interest involves the influence of a Level 2 predictor (e.g., What is the impact of company size on well-being?), (c) it is of interest to determine whether a predictor exerts the same influence at the individual and cluster level (i.e., Is the influence of an individual's workload on well-being the same as the influence of a company's average workload on well-being?), and (d) it is of interest to examine interaction effects (e.g., Does company size moderate the relationship between individual workload and well-being?). These four categories are obviously not exhaustive and do not capture every research question that MLMs are capable of addressing. Nevertheless, we believe that these scenarios are inclusive of many substantive applications of MLM found in the behavioral science literature. Again, it is important to emphasize that different forms of centering may be required to address research questions within the same study, so the centering recommendations below should not be viewed as mutually exclusive decisions.

A Level 1 Predictor Is of Substantive Interest

CWC may be the most appropriate form of centering in situations in which the primary substantive interest involves a Level 1 (i.e., person level) predictor. The rationale for this suggestion follows from the fact that CWC removes all between-cluster variation from the predictor and yields a "pure" estimate of the pooled within-cluster (i.e., Level 1) regression coefficient. Raudenbush and Bryk (2002) summarized this situation, stating that "when an unbiased estimate of β_w [i.e., the Level 1 relationship] is desired, group-mean centering [i.e., CWC] will produce it" (p. 139). The use of CWC is also beneficial in this situation because it yields a more accurate estimate of the slope variance. Again, Raudenbush and Bryk (2002) suggested the use of CWC for this purpose, stating that "we recommend group-mean centering [CWC] to detect and estimate properly the slope heterogeneity" (p. 143).

The substantive focus on a Level 1 predictor and the subsequent use of CWC implies that an individual's relative position within a group is an important determinant of his or her behavior (i.e., the frog pond effect; Davis, 1966). For example, it has been hypothesized that academic self-concept is influenced by one's immediate peers, such that a student may feel less efficacious when surrounded by high-achieving peers (Marsh & Hau, 1987, 2003). Per the occupational stress example, researchers have posited that strain may result when an individual's workload is high relative to others in their immediate workgroup because of perceptions of unfairness (Bliese & Jex, 2002). In situations such as these, it is natural to conceive of the multilevel analysis as a two-step procedure, whereby separate models are first fit to each organization's data, and the resulting regression coefficients are subsequently aggregated in order to obtain an estimate of the pooled within-cluster slope. This type of substantive focus naturally lends itself to CWC (Kreft et al., 1995).

A Level 2 Predictor Is of Substantive Interest

In this scenario, the primary substantive focus is on a Level 2 (i.e., cluster level) predictor variable (e.g., What is the impact of company size on well-being?). The cluster-level variable of interest may be an aggregate of the individual scores within each cluster (e.g., average work hours) or may contain no systematic variation at the individual level (e.g., company size); Susser (1994) referred to these as *contextual* and *integral* variables, respectively. At one extreme, there may no Level 1 predictors in the model, so the centering of Level 2 variables would simply follow the prescribed practice in the OLS literature (e.g., Aiken & West, 1991). However, it is probably more typically the case that both Level 1 and Level 2 variables are included in the model, in which case the Level 1 (e.g., person level) predictors might be viewed as nuisance variables that need to be controlled for. A prototypical example of this type of application is a cluster randomized study where Level 2 units (e.g., organizations, schools) are assigned to participate in either the treatment or control condition. In this case it would be of interest to assess the treatment effect (e.g., a Level 2 dummy variable), controlling for individual differences on a number of Level 1 covariates. We subsequently demonstrate that CGM is ideally suited for this situation.

Returning to the occupational stress scenario, suppose that it was of interest to examine the influence of group size (*SIZE*) on psychological well-being, controlling for individual differences in workload. Although we have not yet discussed models with Level 2 predictors, CGM is the method of choice for assessing the impact of cluster-level variables, controlling for Level 1 covariates. To understand why this is so, we examine the combined model regression equation that results from adding workgroup size to the model, as follows:

$$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(SIZE_j) + \gamma_{10}(HOURS_{cgm}) + [\text{residuals}]. \quad (8)$$

For simplicity, the collection of residual terms from Equation 4 has been omitted and has been replaced by a single term in brackets. Because the Level 1 predictor, $HOURS_{cgm}$, is a composite of within- and between-cluster variation, it is correlated with the Level 2 predictor, *SIZE*. As such, γ_{01} is a partial regression coefficient that reflects the influence of group size, controlling for workload. It is still true that the regression coefficient associated with workload (γ_{10}) is an ambiguous mixture of the Level 1 and Level 2 association between workload and well-being, but this is less of a concern, given that the substantive focus is on the Level 2 covariate and its corresponding regression coefficient (γ_{01}).

It is important to note that CWC does not control for the effects of Level 1 covariates and thus would be inappropriate in this scenario. This conclusion can be reached by substituting $HOURS_{cwc}$ into Equation 8 in place of $HOURS_{cgm}$. Level 1 predictors that are centered at the group mean are uncorrelated with Level 2 variables, so the use of CWC in this situation would result in two orthogonal predictors, and the influence of *SIZE* would be evaluated, independent of work hours. In fact, the regression slope for workgroup size would be the same regardless of whether $HOURS_{cwc}$ was even in the model.

To illustrate the concepts presented thus far, we generated an artificial data set with 300 clusters of 40 cases each. The data were roughly consistent with Figure 1, such that the Level 2 association between workload and well-being was stronger than that at Level 1; both variables had an intraclass correlation of approximately .50. Additionally, the Level 2 variable *SIZE* was negatively correlated with well-being and accounted for a significant proportion of the variation in well-being scores, above and beyond that of workload. Consistent with recommendations from the OLS regression literature, the Level 2 predictor was centered at the grand mean. All models were estimated by using full maximum likelihood, and the analyses were performed with the mixed procedure in SPSS 14 (Peugh & Enders, 2005). The artificial data and SPSS syntax for the example analyses are available from the authors and can also be obtained online.

We first estimated an unconditional model with no predictors at either level of the hierarchy, the results from which are given in Table 2. As seen in the table, the intercept value was $\gamma_{00} = 60.07$, and the intercept variance estimate was $\tau_{00} = 47.36$.

The next section in Table 2 gives the results from two models, one of which included $HOURS_{cgm}$ as a predictor and the other of which included $HOURS_{cwc}$. First, notice that the CGM regression coefficient was slightly larger in absolute value than that of CWC ($\gamma_{10} = -.31$ vs. $-.29$, respectively). Recall from a previous discussion that CWC is preferred when the substantive focus involves a Level 1 variable, because the γ_{10} coefficient is an unbiased estimate of the within-cluster regression. In contrast, the corresponding CGM coefficient is a blend of the Level 1 and Level 2 regressions. The fact that the CGM coefficient was somewhat larger than that of CWC is expected in this case, given that the association between workload and well-being was stronger at Level 2 (i.e., the CGM coefficient is getting

Table 2
Parameter Estimates (PEs) From Example Analysis 1

Model		CGM PE	CGM *SE*	CGM *t*	CWC PE	CWC *SE*	CWC *t*
$WELLBEING_{ij} = \gamma_{00} + u_{0j} + r_{ij}$	γ_{00}	60.074	.403	149.25			
	τ_{00}	47.355	3.969	11.93			
	σ^2	49.956	.653	76.49			
$WELLBEING_{ij} = \gamma_{00} + \gamma_{10}(HOURS_{ij}) + u_{0j}$	γ_{00}	59.927	.353	169.97	60.074	.403	149.25
$+ u_{1j}(HOURS_{ij}) + r_{ij}$	γ_{10}	−.306	.012	−24.54	−.295	.013	−23.06
	τ_{00}	35.762	3.064	11.67	47.492	3.969	11.97
	τ_{11}	.024	.004	6.23	.025	.004	6.31
	τ_{10}	−.593	.084	−7.01	−.606	.095	−6.36
	σ^2	44.474	.589	75.54	44.469	.589	75.51
	γ_{00}	59.878	.337	177.46	60.074	.398	150.93
$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(SIZE_j) + \gamma_{10}(HOURS_{ij})$	γ_{01}	−.165	.029	−5.63	−.096	.036	−2.64
$+ u_{0j} + u_{1j}(HOURS_{ij}) + r_{ij}$	γ_{10}	−.311	.012	−25.00	−.295	.013	−23.06
	τ_{00}	32.624	2.821	11.57	46.418	3.881	11.96
	τ_{11}	.023	.004	6.20	.025	.004	6.31
	τ_{10}	−.581	.081	−7.17	−.601	.094	−6.37
	σ^2	44.486	.589	75.54	44.468	.589	75.51

Note. All significance tests were significant at $p < .01$. Significance tests for variance components are Wald z tests. All residual terms were generated to have a normal distribution. Results from the unconditional model were arbitrarily listed under CGM, although no predictors were in the model. As a result of rounding error in the tabled values, some test statistics do not equal the estimate divided by the standard error. CGM = centering at the grand mean; CWC = centering within cluster; *WELLBEING* = psychological well-being; *HOURS* = workload measured in hours per week; *SIZE* = workgroup size.

"pulled" toward the Level 2 regression line). Second, note that the CWC estimate of the intercept variance was virtually identical to that obtained from the unconditional model, which follows from the definition of β_{0j} as an unadjusted mean. In contrast, the CGM estimate of τ_{00} was substantially lower and reflects the variation in the adjusted outcome means (i.e., the variation in the well-being means, having partialled out the effect of workload). The τ_{10} term is the covariance between the intercepts and slopes, and differs in value due to the fact that CGM and CWC have different definitions of β_{0j}.

The final two models in Table 2 include both workload (a Level 1 variable) and group size (a Level 2 variable) as covariates. First, notice that the CWC estimate of the workload regression coefficient was unaffected by the inclusion of group size and follows from the fact that $HOURS_{cwc}$ is orthogonal to Level 2 variables. As such, the CWC regression coefficients reflect the independent influence of workload and group size as Level 1 and Level 2 predictors, respectively (i.e., workload is not partialled out when considering the influence of group size). Under CWC, the γ_{10} regression coefficient still provides an unbiased estimate of the within-cluster regression, but the model now yields an estimate of the regression of $\bar{y}_{WELLBEING_j}$ on *SIZE*. In contrast, the CGM estimate of the workload regression coefficient changed slightly when group size was added to the model because $HOURS_{cgm}$ is correlated with both Level 1 and Level 2 variables. In this situation, the γ_{01} regression coefficient quantifies the influence of *SIZE*, controlling for individual workload. The γ_{10} regression coefficient still gives a distorted view of the Level 1 regression of well-being on workload, but this is not a concern if the substantive focus is on the Level 2 covariate and the γ_{01} coefficient.

It Is of Interest to Examine a Predictor's Influence at Two Levels

In certain situations it may be of interest to determine whether the association between *X* and *Y* is the same at both levels of the hierarchy. That is, it is of interest to determine whether the person-level regression of *Y* on *X* is the same as the Level 2 regression of $\bar{y}_j$ on $\bar{x}_j$. These models have been referred to as *contextual*, or *compositional*, models in the MLM literature, and are frequently of interest in education and sociology (Blalock, 1984; Raudenbush, 1989b; Raudenbush & Bryk, 2002). Per the occupational stress example, the differential impact of a predictor at two levels can be seen graphically in the top panel of Figure 3, in which the organization-level effect of workload is much stronger, or has a steeper slope, than the corresponding person-level effect; the dotted regression line represents the pooled within-cluster regression of well-being on work hours (i.e., the person-level effect) and the solid line represents the regression of the well-being means on the workload means (i.e., the cluster-level effect).

In order to test whether a predictor has a differential effect at both levels of the hierarchy, we must use the individual scores and the cluster means as predictors in the model. Despite the fact that the Level 2 predictor is an aggregate of the individual scores within each cluster, it is not necessarily true that *X* and $\bar{x}$ share the same meaning or that they measure the same construct (Firebaugh, 1978). For example, in the occupational stress literature, it has been argued that the average work hours within a workgroup may be dictated by externally mandated work requirements, whereas an individual's workload may be driven by a desire to get ahead, a desire to avoid one's family, etc. (Bliese, 2000; Bliese & Halverson, 1996). A similar distinction has been made between *X* and $\bar{x}$ in the public health literature, in which individual and aggregated neighborhood poverty may exert different, independent effects on health outcomes (Schwartz, 1994). We include these two examples in order to highlight the complexities associated with interpreting a Level 2 variable that is an aggregate of individual scores. A more thorough discussion of this issue can be found in a number of different sources (Bliese & Jex, 2002; Chan, 1998; Firebaugh, 1978; Willms, 1986).

Contextual models have a long history in education and sociology, but the interest in *X* as a predictor at two levels also has potential applications in psychology. For example, suppose that it was of interest to examine the influence of anxiety on pain perception. Furthermore, suppose that multiple measurements of both variables were obtained from hospitalized individuals at a number of random intervals throughout the course of a week. This is a somewhat different application from the one we have been using throughout this article, as repeated measurements (i.e., Level 1) are now nested within individuals (i.e., Level 2). Regardless, it is possible to ask whether there is a differential association between anxiety and perceived pain at Level 1 and at Level 2. The Level 1 association between anxiety and pain addresses whether, for any given person, perceived pain is high when anxiety is high (i.e., How do state-like fluctuations in anxiety influence pain ratings?). In contrast, the Level 2 relationship asks whether individuals with high average anxiety scores also have high pain ratings (i.e., How is trait-like variation in anxiety related to perceived pain?). The preceding example is inconsistent with the traditional definition of a contextual model, but it is important to point out that the models described in this section have applications in psychology. We now return to the occupational stress example for the remainder of this section.

In situations in which it is of interest to determine whether the association between *X* and *Y* is the same at both levels of the hierarchy, it is necessary to decompose the predictor into a within- and a between-level component. This is accomplished by using the cluster means, $\bar{x}_{HOURS_j}$, as a predictor in the Level 2 intercept equation. Under CGM, the combined model equation becomes

$$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(\bar{x}_{HOURS_j}) + \gamma_{10}(HOURS_{cgm}) + [\text{residuals}], \quad (9)$$

and the corresponding CWC model is

$$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(\bar{x}_{HOURS_j}) + \gamma_{10}(HOURS_{cwc}) + [\text{residuals}]. \quad (10)$$

Equations 9 and 10 are interesting because they represent one of the few situations in which CGM and CWC produce equivalent parameter estimates (Kreft et al., 1995), so either form of centering is appropriate. In this context, equivalence means that the parameter estimates from CGM can be algebraically equated with those of CWC, even though the actual values of the two sets of coefficients differ.

The fact that both CGM and CWC can be used to test whether a differential association exists between *X* and *Y* at Level 1 and Level 2 stems from the following algebraic relationship given by Kreft et al. (1995): $\gamma_{01}^{cgm} = \gamma_{01}^{cwc} - \gamma_{10}^{cwc}$. The equivalence of these parameters can be understood by examining the regression coefficients in Equations 9 and 10. Level 1 predictors that are centered at the group mean are uncorrelated with Level 2 variables, so the CWC model in Equation 10 is comprised of two orthogonal predictors, and γ_{10} and γ_{01} quantify the independent association between workload and well-being at Level 1 and Level 2, respectively. In order to determine whether a differential association exists between *X* and *Y* at the two levels of the hierarchy (i.e., a contextual effect exists), it is necessary to test whether the two coefficients are statistically different from one another. In contrast, the predictor variables in the CGM model shown in Equation 9 are highly correlated (because $HOURS_{cgm}$ is, in part, comprised of $\bar{x}_{HOURS_j}$), so the γ coefficients are partial regression slopes that quantify the effect of workload at one level, controlling for the influence of workload at the other level. In this situation, a differential relationship between *X* and *Y* at the two levels would be evident if the cluster means are associated with *Y*, after controlling for the effect of *X* at Level 1. Said differently, if the magnitude of the association between *X* and *Y* is identical at both levels, then the cluster means would provide no additional explanatory power, and γ_{01} would equal zero. Most multilevel software packages allow the user to specify custom contrasts, so testing the difference between the Level 1 and Level 2 regression slopes is straightforward under CWC. Nevertheless, the use of CGM is arguably easier in this context, given that the user need only examine the significance test for γ_{01}.

Kreft et al. (1995) also illustrated that CGM and CWC yield identical estimates of the Level 1 association between *X* and *Y*, such that $\gamma_{10}^{cgm} = \gamma_{10}^{cwc}$. At first glance, this algebraic relationship appears to be inconsistent with the earlier conclusion that CGM yields a biased estimate of the Level 1 regression. However, adding the cluster means as a predictor in Equation 9 serves to partial out the Level 2 influence of workload, resulting in an unbiased estimate of the relationship between individual workload and well-being.

Finally, Kreft et al. (1995) noted that the CGM and CWC intercepts can be algebraically equated such that $\gamma_{00}^{cgm} - \gamma_{10}^{cgm}\bar{x} = \gamma_{00}^{cwc}$. It is worth noting that the CGM and CWC regression coefficients in Equations 9 and 10 are equivalent regardless of how one chooses to model the random effects. However, CGM and CWC variance components are only equivalent when the intercepts, but not the slopes, randomly vary across clusters. The equivalence of contextual model parameters was explored in detail by Kreft et al. (1995), so readers who are interested in a more rigorous treatment of this topic are encouraged to consult their earlier work.[3]

The second set of illustrative analyses involved the models from Equations 9 and 10, and used the artificial data set described in the previous example. Parameter estimates from these analyses are given in Table 3.

We have previously shown that both CGM and CWC can be used to test whether a differential association exists between *X* and *Y* at Level 1 and at Level 2. Under CWC, γ_{10} and γ_{01} quantify the independent association between workload and well-being at Level 1 and Level 2, respectively, so the primary focus of this analysis involves testing whether $\gamma_{10} = \gamma_{01}$. As shown in Table 3, the difference between γ_{10} and γ_{01} was approximately $-.314$, and a custom hypothesis test indicated that this difference was statistically different from zero ($t = -7.39$, $p < .01$). Under CGM, the γ_{01} regression coefficient quantifies the additional explanatory power of the cluster means, controlling for influence of workload at Level 1, so the primary focus of the CGM analysis involves testing whether γ_{01} differs from zero; as shown in Table 3, the γ_{01} coefficient was $-.315$, and was statistically significant. Notice that the CGM estimate of γ_{01} was nearly identical to the difference between γ_{10} and γ_{01} under CWC and follows from the fact that the CGM and CWC models shown in Equations 9 and 10 are equivalent (Kreft et al., 1995).

Two additional points should be made about the parameter estimates in Table 3. First, notice that both models produced the same estimate of the Level 1 association between workload and well-being—this follows from Kreft et al. (1995). As noted previously, the inclusion of the cluster means in the CGM model serves to partial out the

[3] Because of a lack of precision in the estimation process, the algebraic identities outlined by Kreft et al. (1995) may not hold exactly when substituting the estimated parameter values into the equations outlined in this section.

Table 3
Parameter Estimates (PEs) From Example Analysis 2

Model		CGM PE	CGM *SE*	CGM *t*	CWC PE	CWC *SE*	CWC *t*
$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(\bar{x}_{HOURS_j}) + \gamma_{10}(HOURS_{ij})$	γ_{00}	60.064	.332	180.77	60.074	.320	187.90
$+ u_{0j} + u_{1j}(HOURS_{ij}) + r_{ij}$	γ_{01}	−.315	.043	−7.30	−.608	.040	−15.04
	γ_{10}	−.294	.013	−23.44	−.294	.013	−22.99
	τ_{00}	31.475	2.701	11.65	29.552	2.504	75.52
	τ_{11}	.024	.004	6.21	.025	.004	6.34
	τ_{10}	−.584	.080	−7.28	−.595	.079	−7.58
	σ^2	44.465	.589	75.55	44.464	.589	75.52

Note. All significance tests were significant at $p < .01$. Significance tests for variance components are Wald z tests. All residual terms were generated to have a normal distribution. As a result of rounding error in the tabled values, some test statistics do not equal the estimate divided by the standard error. CGM = centering at the grand mean; CWC = centering within cluster; *WELLBEING* = psychological well-being; *HOURS* = workload measured in hours per week.

between-cluster variation in workload, resulting in an unbiased estimate of the Level 1 regression slope. Finally, note that the CGM and CWC estimates of the intercept variance (τ_{00}) were quite similar in these two models. Recall from an earlier discussion that the CGM estimate of τ_{00} quantifies variation in the well-being means after adjusting for mean differences in workload. The addition of $\bar{x}_{HOURS_j}$ to the CWC model effectively removed the between-cluster variation in well-being that was attributable to workload, so the CGM and CWC estimates of τ_{00} now have the same interpretation.

An Interaction Effect Is of Substantive Interest

Given the important role that centering plays in OLS regression models with interaction terms (Aiken & West, 1991), this issue has received surprisingly little attention in the MLM literature—popular MLM textbooks are virtually devoid of the topic. Hofmann and Gavin (1998) underscored the importance of this issue, stating that “When moving to models that include cross-level interactions, the differences between grand mean and group mean centering become even more critical” (p. 631).

MLM analyses allow for the estimation of within-level and cross-level interactions. We focus primarily on cross-level interactions in this section, briefly addressing within-level interactions as well. A cross-level interaction occurs when a Level 2 variable moderates the magnitude of a Level 1 relationship. Returning to the occupational stress example, suppose that group size (an organization-level variable) moderated the relationship between workload and well-being (the person-level regression), such that the association between workload and well-being was weaker in small workgroups. A cross-level interaction effect is modeled by adding a Level 2 covariate to the slope formula shown in Equation 3. Under CGM, the combined model regression equation is as follows:

$$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(SIZE_j) + \gamma_{10}(HOURS_{cgm}) + \gamma_{11}(SIZE_j)(HOURS_{cgm}) + [\text{residuals}] \quad (11)$$

Again, the collection of residual terms from Equation 4 is represented by a single term in brackets. The γ_{11} regression coefficient is of particular interest in this context because it quantifies the moderating influence that the group size has on magnitude of the Level 1 association between workload and well-being.

The impact of centering on cross-level interactions was raised in a classic article by Cronbach and Webb (1975) that reexamined an Aptitude × Treatment interaction (ATI) in a study of mathematics achievement. A previous analysis had incorrectly concluded that an ATI was present, when this effect was actually due to an interaction at the classroom level (i.e., Level 2). Specifically, the association between the aptitude and achievement means differed for treatment and control classrooms, but a failure to separate out the within-level (i.e., individual) and between-level (i.e., classroom) relationship between X and Y led to the incorrect conclusion concerning the ATI. Cronbach and Webb hypothesized that a true ATI would be evidenced by a cross-level interaction, such that the within-class (i.e., Level 1) relationship between aptitude and achievement would be attenuated by participation in the treatment (i.e., a Level 2 variable). However, appropriately disentangling the within- and between-cluster relationship revealed no such effect.

Hofmann and Gavin (1998) noted a problem similar to that of Cronbach and Webb (1975) and used artificial data to demonstrate that CGM can produce a significant cross-level interaction effect, when no such effect exists in the population. Again, the problem with CGM stems from the fact that $HOURS_{cgm}$ is a composite of within- and between-cluster variation. This idea makes it possible to express the cross-level interaction using the following formula: $\gamma_{11}(SIZE_j)(HOURS_{within} + HOURS_{between})$. With some additional algebra, the formula becomes $\gamma_{11}(SIZE_j) \times (HOURS_{within}) + \gamma_{11}(SIZE_j)(HOURS_{between})$. The previous expression illustrates the important point that γ_{11} is potentially influenced by two qualitatively different interaction effects: the interaction between *SIZE* and the Level 1 variation in workload scores and the Level 2 interaction between *SIZE* and the workload means. As such, the CGM estimate of γ_{11} suffers from the same problem that was encountered when estimating the person-level association between X and Y, namely that it is an uninterpretable mixture of two different influences on the outcome.

Now reconsider the previous conceptual formula after substituting $HOURS_{cwc}$ in place of $HOURS_{cgm}$. Group mean centering removes all between-cluster variation in X, so the formula simplifies to $\gamma_{11}(SIZE_j)(HOURS_{within})$. At a conceptual level, CWC yields a pure estimate of the moderating influence that a Level 2 predictor exerts on the Level 1 association between X and Y and cannot be distorted by the presence of an interaction that involves the cluster means of X. It is for this reason that Hofmann and Gavin (1998) and Raudenbush (1989a, 1989b) recommended using CWC when cross-level interactions are of substantive interest.

The potential confounding that occurs under CGM also implies that CWC is appropriate when the substantive focus involves an interaction between a pair of Level 1 variables. At a conceptual level, a cross-level interaction and an interaction involving a pair of Level 1 variables both require an accurate estimate of the Level 1 slope because this coefficient is being moderated by another predictor. The need for an unbiased estimate of the Level 1 association makes CWC a natural choice for examining these types of interaction effects.

Although the model in Equation 11 does lead to the possibility of a spurious cross-level interaction effect (Hofmann & Gavin, 1998), it ends up that both CGM and CWC can be used to appropriately distinguish a cross-level interaction from an interaction involving the Level 2 moderator and the cluster means (i.e., the two interaction terms that are confounded in Equation 11). Returning to the occupational stress example, we consider the following combined model equation:

$$Y_{ij} = \gamma_{00} + \gamma_{01}(\bar{x}_{HOURS_j}) + \gamma_{02}(SIZE_j) + \gamma_{03}(\bar{x}_{HOURS_j})(SIZE_j)$$

$$+ \gamma_{10}(HOURS_{ij}) + \gamma_{11}(HOURS_{ij})(SIZE_j) + [residuals]. \quad (12)$$

The model outlined above includes a main effect for the Level 1 predictor (i.e., γ_{10}), main effects for the Level 2 predictor and the cluster means of the Level 1 predictor (i.e., γ_{02} and γ_{01}, respectively), a cross-level interaction (i.e., γ_{11}), and an interaction involving the Level 2 predictor and the cluster means (i.e., γ_{03}). The notation in Equation 12 is somewhat generic because either $HOURS_{cgm}$ or $HOURS_{cwc}$ can be used in place of $HOURS_{ij}$. The collection of residual terms in the bracket is identical to that in Equation 4.

It is interesting to note that the model in Equation 12 represents another special case in which CGM and CWC yield equivalent parameter estimates; as before, equivalence means that the parameters in the CGM model can be algebraically equated to those in the CWC model. In the interest of space, the analytic details related to model equivalence are given in Appendix A, and we instead focus on the substantive interpretation of the interaction coefficients under CGM and CWC. The interpretation of Equation 12 relies on two previously established points: CGM yields Level 1 scores that are correlated with both Level 1 and Level 2 variables, whereas CWC yields Level 1 scores that are orthogonal to all Level 2 variables. Because $HOURS_{cwc}$ and $\bar{x}_{HOURS_j}$ are orthogonal, it follows that the γ_{11} and γ_{03} coefficients reflect the independent influence of the cross-level interaction and the interaction involving *SIZE* and the workload cluster means, respectively. In contrast, $HOURS_{cgm}$ and $\bar{x}_{HOURS_j}$ are correlated, so γ_{03} is interpreted as the unique influence of the interaction involving *SIZE* and the cluster means, above and beyond that of the cross-level interaction (i.e., the effect of the Level 2 interaction, having partialled out the cross-level interaction). Because the CGM interaction effects are interpreted as partial regression coefficients, it also follows that γ_{11} is a pure estimate of the cross-level interaction, and is no longer confounded with the Level 2 interaction, because it has been appropriately partialled out. In fact, the analytic details given in Appendix A demonstrate that CGM and CWC yield identical estimates of the cross-level interaction in this case.

It is unclear whether the two interaction effects actually occur simultaneously in practice, but the model outlined in Equation 12 allows one to appropriately disentangle the two effects by using either CGM or CWC. The model in Equation 12 is also interesting because it is a second case in which CGM and CWC yield equivalent regression coefficients; CGM and CWC also yield equivalent estimates of the contextual model parameters in Equations 9 and 10 (Kreft et al., 1995). The equivalence of CGM and CWC has not been widely studied, so it is unclear whether more general conditions of equivalence can be established.

To illustrate the effect of centering on interaction effects, we again generated an artificial data set with 300 clusters of 40 cases each. The data were generated such that the outcome variable scores were a function of the Level 2 interaction involving *SIZE* and $\bar{x}_{HOURS_j}$, and all other terms in the model were set to zero.

We began by estimating a standard cross-level interaction model such as that given in Equation 11. As seen in Table 4, CGM produced a significant cross-level interaction, even though no such effect existed in the data. This finding is similar to that reported by Hofmann and Gavin (1998) and demonstrates the pitfalls associated with using CGM to investigate cross-level interaction effects. In contrast, CWC appropriately disentangled the two interaction effects and produced a nonsignificant cross-level interaction.

The interaction model proposed in Equation 12 was used for the second analysis and was estimated by using both

Table 4
Parameter Estimates (PEs) From Example Analysis 3

Model		CGM PE	CGM *SE*	CGM *t*	CWC PE	CWC *SE*	CWC *t*
$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(SIZE_j) + \gamma_{10}(HOURS_{ij}) + \gamma_{11}(HOURS_{ij})(SIZE_j) + u_{0j} + u_{1j}(HOURS_{ij}) + r_{ij}$	γ_{00}	59.394	0.403	**147.34**	59.389	0.415	**143.12**
	γ_{01}	−0.028	0.040	−0.70	−0.022	0.041	0.55
	γ_{10}	0.013	0.010	1.25	0.012	0.010	1.17
	γ_{11}	0.003	0.001	**3.02**	0.001	0.001	0.97
	τ_{00}	48.818	0.647	**75.51**	50.433	4.217	**11.96**
	τ_{11}	0.007	0.003	**2.61**	0.007	0.003	**2.56**
	τ_{10}	0.094	0.071	1.33	0.052	0.075	0.69
	σ^2	48.818	0.647	**75.51**	48.788	0.646	**75.50**
$WELLBEING_{ij} = \gamma_{00} + \gamma_{01}(\bar{x}_{HOURS_j}) + \gamma_{02}(SIZE_j) + \gamma_{03}(\bar{x}_{HOURS_j})(SIZE_j) + \gamma_{10}(HOURS_{ij}) + \gamma_{11}(HOURS_{ij})(SIZE_j) + u_{0j} + u_{1j}(HOURS_{ij}) + r_{ij}$	γ_{00}	59.967	0.274	**218.76**	59.976	0.274	**219.12**
	γ_{01}	0.037	0.039	0.96	0.049	0.038	1.31
	γ_{02}	−0.022	0.027	−0.81	−0.023	0.027	−0.84
	γ_{03}	0.069	0.004	**18.89**	0.070	0.003	**19.99**
	γ_{10}	0.012	0.010	1.19	0.012	0.010	1.18
	γ_{11}	0.001	0.001	0.98	0.001	0.001	0.97
	τ_{00}	20.834	1.818	**11.46**	20.996	1.814	**11.57**
	τ_{11}	0.005	0.002	**2.16**	0.007	0.003	**2.55**
	τ_{10}	0.093	0.047	**1.99**	0.072	0.049	1.45
	σ^2	48.859	0.647	**75.48**	48.790	0.646	**75.50**

Note. Bold typeface denotes statistical significance at $p < .05$. Significance tests for variance components are Wald *z* tests. All residual terms were generated to have a normal distribution. As a result of rounding error in the tabled values, some test statistics do not equal the estimate divided by the standard error. CGM = centering at the grand mean; CWC = centering within cluster; *WELLBEING* = psychological well-being; *SIZE* = workgroup size; *HOURS* = workload measured in hours per week.

CGM and CWC. Most important, the two interaction effects were appropriately disentangled, such that the cross-level interaction was nonsignificant, and the Level 2 interaction involving *SIZE* and $\bar{x}_{HOURS_j}$ was statistically significant.[4] As outlined in Appendix A, CGM and CWC now yield identical estimates of the cross-level interaction. More generally, the parameter estimates given in the bottom portion of Table 4 are consistent with the algebraic identities given in Appendix A (e.g., $\gamma_{03}^{cgm} = \gamma_{03}^{cwc} - \gamma_{11}^{cwc} = .069 = .070 - .001$). Although the two sets of estimates are algebraically equivalent in this situation, we believe that the CWC model is more useful, as it leads to a more natural interpretation of the two interaction effects (i.e., the interaction coefficients are independent under CWC, whereas the CGM coefficients reflect the influence of one interaction, partialling out the other).

Centering Binary Level 1 Predictors

Having established some centering recommendations for continuous Level 1 predictors, we now briefly discuss issues of centering with binary predictors (e.g., gender). Although it may seem unnatural to center nominal predictors, the previous concepts also extend to dummy and effect code variables that appear in the Level 1 model. We provide a nontechnical discussion of centering in this section, but a more detailed algebraic presentation is given in Appendix B.

First, consider a Level 1 dummy variable (i.e., $x_{ij} = 0, 1$) with n_{0j} cases in the base group (i.e., $x_{ij} = 0$) and n_{1j} cases in the comparison group (i.e., $x_{ij} = 1$). Returning to the occupational stress example, suppose that it was of interest to include a dummy code for gender in the Level 1 model. In this situation, the Level 1 intercept (i.e., β_{0j}) would be interpreted as the male mean in cluster *j* (i.e., the expected value when *FEMALE* = 0), and τ_{00} would quantify the variation in the male averages across the *j* clusters. Alternatively, gender could be represented as an effect code variable (i.e., male = −1, female = 1), in which case β_{0j} would be interpreted as the unweighted mean of *Y* in cluster *j*, and τ_{00} would quantify the variation in the unweighted means. The interpretation of the Level 1 intercept in the previous two examples is consistent with the interpretation of dummy and effect codes in OLS regression (Cohen, Cohen, West, & Aiken, 2003).

Centering a dummy variable around the grand mean fundamentally changes the interpretation of the MLM intercept but does so in a way that is consistent with the continuous case. Specifically, CGM yields a Level 1 intercept that is

[4] CGM does not yield a direct significance test of the Level 2 interaction involving *SIZE* and $\bar{x}_{HOURS_j}$. Rather, the CGM coefficient quantifies the difference between the Level 2 and cross-level interactions.

interpreted as the adjusted outcome mean for cluster *j*. In the continuous case, the outcome mean was adjusted for mean differences that exist among clusters on the covariate *X* (see Equation 6). The mean of a binary dummy variable is simply the proportion of cases in the comparison group (e.g., *FEMALE* = 1), so the CGM intercept for cluster *j* is adjusted for differences in the proportion of comparison group cases across clusters. That is, β_{0j} is the cluster mean that would result had the proportion of females been identical across workgroups. Consistent with the continuous case, τ_{00} quantifies the variation in the adjusted outcome means. It is interesting to note that applying CGM to a binary effect code variable yields an identical interpretation, so the choice of coding has no bearing on the interpretation of the intercept or on the intercept variance, provided that the data are centered. Algebraic support for this conclusion is given in Appendix B.

Applying CWC to a dummy or effect code variable also yields the same interpretation as it did in the continuous case. Specifically, applying CWC to a Level 1 dummy variable (e.g., deviating each person's gender value around the proportion of females in cluster *j*) yields a Level 1 intercept that is interpreted as the unadjusted (i.e., weighted) mean for cluster *j*. In fact, the CWC intercept is interpreted as an unadjusted mean, regardless of whether one chooses to use dummy or effect coding in the Level 1 equation, so the choice of coding scheme is completely arbitrary. Again, the algebraic details supporting this conclusion are given in Appendix B.

In summary, the interpretation of the MLM intercept is unaffected by the use of dummy or effect codes and depends solely on the method of centering used. Under CWC, β_{0j} is interpreted as the unadjusted (i.e., weighted) mean for cluster *j*, and τ_{00} quantifies the variation in the outcome variable means. Under CGM, the dependent variable means are adjusted for differences in the proportion of comparison group cases (i.e., $x_{ij} = 1$) across clusters, so β_{0j} is interpreted as the adjusted mean for cluster *j*. As such, τ_{00} quantifies the variation in the adjusted means. Again, note that these interpretations are identical to the continuous case.

The coding scheme at Level 1 does affect the within-cluster regression coefficient. Consistent with dummy coding in OLS regression, β_{1j} equals the group mean difference within cluster *j* and is equal to half the mean difference when effect codes are applied (see Cohen et al., 2003). It also follows that the slope variance estimate (i.e., τ_{11}) will differ across coding schemes, as will the covariance between the intercepts and slopes (i.e., τ_{01}). Specifically, the slope variance under dummy coding will be exactly four times larger than the variance under effect coding. However, the significance tests for the variance components will be identical under dummy and effect coding, so the choice of coding scheme makes very little difference.

As a cautionary note, we encourage readers to carefully consider the impact of leaving a dummy or effect code variable uncentered in the Level 1 model because doing so may produce an awkward interpretation of certain model parameters. For example, suppose that Equation 5 was expanded to include an uncentered dummy code for gender (i.e., male = 0, female = 1). Because workload was grand mean centered, the Level 1 intercept would be interpreted as the male mean, adjusted for cluster differences in workload. By extension, τ_{00} would then quantify the variation in the adjusted outcome means for males; had both variables been grand mean centered, τ_{00} would have quantified the variation in the well-being means, having partialled out both gender and workload. In summary, the use of a binary variable in the Level 1 equation does not change our centering recommendations, so researchers should carefully consider the impact of centering, regardless of whether the predictor variable is continuous or binary.

Centering Level 2 Predictors

A brief discussion of centering Level 2 predictors is warranted before concluding. Centering at Level 2 is typically far less complex than the centering decisions required at Level 1 because it is only necessary to choose between the raw metric and CGM; CWC is not an option because scores on a Level 2 variable are constant within each cluster. In general, centering decisions at Level 2 can be based on recommendations from the OLS regression literature (Aiken & West, 1991; Cohen et al., 2003). For example, if the Level 2 equation contains only first-order terms, choosing between the raw metric and CGM will only affect the intercept (γ_{00}). Consistent with standard practice, CGM is preferred when higher order terms are added to the Level 2 model (e.g., an interaction between a pair of Level 2 variables, or a quadratic effect in the Level 2 model). CGM generally provides a convenient option for centering Level 2 variables, but it may also be useful to leave a variable in its raw metric (e.g., a dummy or effect coded variable).

Summary and Conclusions

Psychological constructs are frequently expressed on arbitrary metrics that lack a clearly interpretable or meaningful zero value, and centering is frequently used to establish a useful zero point. However, the centering of Level 1 predictors in MLM analyses is complex, and confusion continues to exist over the appropriate use of CGM and CWC. The purpose of this article was to explicate the vital role that centering plays in defining MLM parameter estimates and to provide researchers with rules of thumb to guide centering decisions in their own research.

The decision to use CGM or CWC cannot be based on statistical evidence, but depends heavily on one's substantive research questions. This issue is clearly complex, and no single model may sufficiently address an entire set of substantive research questions. In fact, it seems quite reasonable to use CGM and CWC within the context of a single study. It is possible that one research question requires the use of CGM, whereas a different question is best addressed with CWC. Even when the substantive question calls for CGM, Raudenbush and Bryk (2002) noted that CWC may provide the most accurate estimates of the slope variance due to the negative bias that can sometimes be present in the CGM estimate of this parameter.

We proposed some straightforward rules of thumb that we hope will be of use to substantive researchers and methodologists alike. These guidelines are as follows: (a) CWC is appropriate if the Level 1 association between *X* and *Y* is of substantive interest; (b) CGM is appropriate when one is primarily interested in a Level 2 predictor and wants to control for Level 1 covariates; (c) either CGM or CWC can be used to examine the differential influence of a variable at Level 1 and Level 2; and (d) CWC is preferable for examining cross-level interactions and interactions that involve a pair of Level 1 variables, and CGM is appropriate for interactions between Level 2 variables.

The thoughtless application of such guidelines is always problematic, and we encourage readers to use our suggestions as a springboard for examining the linkage between their substantive research questions and the form of centering that will best address those questions. In addition, we strongly recommend that centering decisions are reported in published articles and that authors provide a brief rationale for their analytic choices.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Balsam, K. F., Beauchaine, T. P., Mickey, R. M., & Rothblum, E. D. (2005). Mental health of lesbian, gay, bisexual, and heterosexual siblings: Effects of gender, sexual orientation, and family. *Journal of Abnormal Psychology, 114,* 471–476.

Biesanz, J. C., Deeb-Sossa, N., Aubrecht, A. M., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9,* 30–52.

Blalock, H. M. (1984). Contextual-effects models: Theoretical and methodological issues. *Annual Review of Sociology, 10,* 353–372.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61,* 27–41.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco: Jossey-Bass.

Bliese, P. D., & Halverson, R. R. (1996). Individual and nomothetic models of job stress: An examination of work hours, cohesion, and well-being. *Journal of Applied Social Psychology, 26,* 1171–1189.

Bliese, P. D., & Jex, S. M. (2002). Incorporating a multilevel perspective into occupational stress research: Theoretical, methodological, and practical implications. *Journal of Occupational Health Psychology, 7,* 265–276.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis. A typology of compositional models. *Journal of Applied Psychology, 83,* 234–246.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cronbach, L. J., & Webb, N. (1975). Between- and within-class effects in a reported aptitude-by-treatment interaction: Reanalysis of a study by G. L. Anderson. *Journal of Educational Psychology, 67,* 717–724.

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology, 72,* 17–31.

Firebaugh, G. (1978). A rule for inferring individual-level relationships from aggregate data. *American Sociological Review, 43,* 557–572.

Griffin, M. A. (2001). Dispositions and work reactions: A multilevel approach. *Journal of Applied Psychology, 86,* 1142–1151.

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

Kallestad, J. H., & Olweus, D. (2003, October 1). Predicting teachers' and schools' implementation of the Olweus Bullying Prevention Program: A multilevel study. *Prevention and Treatment, 6.*

Kreft, I. (1995). The effects of centering in multilevel analysis: Is the public school the loser or the winner? A new analysis of an old question. *Multilevel Modelling Newsletter, 7*(3), 5–8.

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.

Longford, N. (1989a). Contextual effects and group means. *Multilevel Modelling Newsletter, 1*(3), 5, 11.

Longford, N. (1989b). To center or not to center. *Multilevel Modelling Newsletter, 1*(3), 7, 11.

Marsh, H. W., & Hau, K.-T. (1987). Big-fish-little-pond effect on academic self concept. *Journal of Educational Psychology, 79,* 280–295.

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self concept. *American Psychologist, 58,* 364–376.

Montague, M., Enders, C., Dietz, S., Morrison, W., & Dixon, J. (in press). A longitudinal study of depressive symptomology and self-concept in adolescents. *Journal of Special Education.*

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event and interval contingent data in social and personality psychology research. *Personality & Social Psychology Bulletin, 27,* 771–785.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review, 30,* 66–85.

Peugh, J. L., & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement, 65,* 717–741.

Plewis, I. (1989). Comment on "centering" predictors in multilevel analysis. *Multilevel Modelling Newsletter, 1*(3), 6, 11.

Raudenbush, S. W. (1989a). "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter, 1*(2), 10–12.

Raudenbush, S. W. (1989b). A response to Longford and Plewis. *Multilevel Modelling Newsletter, 1*(3), 8–11.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Schwartz, S. (1994). The fallacy of the ecological fallacy: The potential misuse of a concept and the consequences. *American Journal of Public Health, 84,* 819–824.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Stevens, J. (1999). *Intermediate statistics: A modern approach* (2nd ed.). Mahwah, NJ: Erlbaum.

Susser, M. (1994). The logic in ecological: The logic of analysis. *American Journal of Public Health, 85,* 825–829.

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73,* 914–923.

Willms, J. D. (1986). Social class segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review, 51,* 224–241.

Wu, Y.-W. B., & Wooldridge, P. J. (2005). The impact of centering first-level predictors on individual and contextual effects in multilevel data analysis. *Nursing Research, 54,* 212–216.

Zohar, D., & Luria, G. (2005). A multilevel model of safety climate: Cross-level relationships between organization and group-level climates. *Journal of Abnormal Psychology, 90,* 616–628.

Appendix A

The Equivalence of CGM and CWC in Equation 16

In the equations below, a single asterisk is used to denote CGM parameters, and a double asterisk is used for CWC. Level 1 and Level 2 predictors are denoted X and W, respectively. Following Kreft et al. (1995), the equivalence of CGM and CWC is demonstrated by finding a solution to Equation A1.

$$\begin{aligned}\gamma^*_{00} + \gamma^*_{01}(\bar{x}_j) + \gamma^*_{02}(W_j) + \gamma^*_{03}(\bar{x}_j)(W_j) + \gamma^*_{10}(X_{ij} - \bar{x}) \\ + \gamma^*_{11}(X_{ij} - \bar{x})(W_j) = \gamma^{**}_{00} + \gamma^{**}_{01}(\bar{x}_j) + \gamma^{**}_{02}(W_j) \\ + \gamma^{**}_{03}(\bar{x}_j)(W_j) + \gamma^{**}_{10}(X_{ij} - \bar{x}_j) + \gamma^{**}_{11}(X_{ij} - \bar{x}_j)(W_j)\end{aligned} \tag{A1}$$

Equation A1 can be further expanded as follows:

$$\begin{aligned}\gamma^*_{00} + \gamma^*_{01}(\bar{x}_j) + \gamma^*_{02}(W_j) + \gamma^*_{03}(\bar{x}_j)(W_j) + \gamma^*_{10}(X_{ij}) \\ - \gamma^*_{10}(\bar{x}) + \gamma^*_{11}(X_{ij}) - \gamma^*_{11}(\bar{x})(W_j) = \gamma^{**}_{00} + \gamma^{**}_{01}(\bar{x}_j) \\ + \gamma^{**}_{02}(W_j) + \gamma^{**}_{03}(\bar{x}_j)(W_j) + \gamma^{**}_{10}(X_{ij}) - \gamma^{**}_{10}(\bar{x}_j) \\ + \gamma^{**}_{11}(X_{ij}) - \gamma^{**}_{11}(\bar{x}_j)(W_j)\end{aligned} \tag{2}$$

Collecting like terms from both sides of Equation A2 yields the following solution:

$$\gamma^*_{00} - \gamma^*_{10}(\bar{x}) = \gamma^{**}_{00} \tag{A3}$$

$$\gamma^*_{01} = \gamma^{**}_{01} - \gamma^{**}_{10} \tag{A4}$$

$$\gamma^*_{02} - \gamma^*_{11}(\bar{x}) = \gamma^{**}_{02} \tag{A5}$$

$$\gamma^*_{03} = \gamma^{**}_{03} - \gamma^{**}_{11} \tag{A6}$$

$$\gamma^*_{10} = \gamma^{**}_{10} \tag{A7}$$

$$\gamma^*_{11} = \gamma^{**}_{11}. \tag{A8}$$

(*Appendixes continue*)

Appendix B

Algebraic Details Associated With Centering Binary Level 1 Predictors

To illustrate the use of CGM, consider a dummy variable (i.e., $x_{ij} = 0, 1$) where the total sample N is comprised of n_0 cases in the base group (i.e., $x_{ij} = 0$) and n_1 cases in the comparison group (i.e., $x_{ij} = 1$). The grand mean for a binary dummy variable is the total proportion of cases in the comparison group, denoted p_1 as follows:

$$\bar{x} = \frac{\Sigma x_{ij}}{N} = \frac{n_0(0) + n_1(1)}{N} = \frac{n_1}{N} = p_1. \quad \text{(B1)}$$

As such, X_{cgm} can be written as $x_{ij} - \bar{x} = x_{ij} - p_1$. Simplifying this expression (i.e., substituting $x_{ij} = 0$ or 1 into the equation) yields centered scores of $-p_1$ and p_0 for the base and comparison group cases, respectively, where p_0 is the total proportion of base group cases.

Under a dummy code scheme, Equation 6 can be rewritten as

$$\begin{aligned}\beta_{0j} &= \bar{y}_j - \beta_{1j}(\bar{x}_j - \bar{x}) = \bar{y}_j - \beta_{1j}(p_{1j} - p_1) \\ &= \bar{y}_j - (\bar{y}_{0j} - \bar{y}_{1j})(p_{1j} - p_1), \quad \text{(B2)}\end{aligned}$$

where β_{1j} equals the mean difference within cluster j, and p_{1j} is the proportion of comparison group cases in cluster j. Equation B2 shows that β_{0j} is equal to the outcome variable mean for cluster j, minus an adjustment term that corrects for differences in the proportion of comparison group cases across clusters.

Applying CGM to an effect code variable in the Level 1 equation yields the same result as the dummy variable case. To illustrate, consider an effect code variable (i.e., $x_{ij} = -1, 1$) where the total sample N is comprised of n_0 cases in the base group (i.e., $x_{ij} = -1$) and n_1 cases in the comparison group (i.e., $x_{ij} = 1$). The grand mean for a binary effect code variable is the difference between the proportion of base and comparison group cases, as follows:

$$\bar{x} = \frac{\Sigma x_{ij}}{N} = \frac{n_0(-1) + n_1(1)}{N} = \frac{-n_0 + n_1}{N} = p_1 - p_0. \quad \text{(B3)}$$

X_{cgm} is now $x_{ij} - \bar{x}_j = x_{ij} - (p_1 - p_0)$, and the centered scores become $-2p_1$ and $2p_0$ for the base and comparison group cases, respectively (i.e., after substituting $x_{ij} = -1$ or 1 into the expression for X_{cgm}).

When using effect codes, β_{1j} equals half the mean difference in cluster j, so Equation 6 can be written as follows

$$\begin{aligned}\beta_{0j} &= \bar{y}_j - \beta_{1j}(\bar{x}_j - \bar{x}) \\ &= \bar{y}_j - (.5)(\bar{y}_{0j} - \bar{y}_{1j})(p_{1j} - p_{0j} - p_1 + p_0), \quad \text{(B4)}\end{aligned}$$

where $\bar{x}_j$ equals $p_{1j} - p_{0j}$. Consistent with a dummy code scheme, applying CGM to a binary effect code variable yields a β_{0j} value that is interpreted as an adjusted mean for cluster j. Although it may not be immediately obvious, the portion of the adjustment that involves the proportion of base and comparison group cases will be twice as large as the corresponding term in Equation B2—intuitively, this follows from the fact that the centered scores are twice as large under effect coding. This is compensated for by the fact that effect coding produces a β_{1j} value that is exactly half as large as it was under dummy coding and results in identical interpretations of β_{0j} and τ_{00}.

The algebraic details associated with CWC are as follows. Consider a dummy variable (i.e., $x_{ij} = 0, 1$) with n_{0j} cases in the base group (i.e., $x_{ij} = 0$) and n_{1j} cases in the comparison group (i.e., $x_{ij} = 1$). Under a dummy coding scheme, $\bar{x}_j$ is equal to the proportion of comparison group cases within cluster j (i.e., $n_{1j}/n_j = p_{1j}$), so X_{cwc} can be written as $x_{ij} - \bar{x}_j = x_{ij} - p_{1j}$. Substituting $x_{ij} = 0$ and 1 into the expression for X_{cwc} yields a centered score of $-p_{1j}$ for the base group cases (i.e., $x_{ij} = 0$), and a code of $1 - p_{1j} = p_{0j}$ for the comparison group cases (i.e., $x_{ij} = 1$). Equation B5 illustrates that the centered dummy code values have a mean of zero within each cluster, which is consistent with the continuous case.

$$\begin{aligned}\bar{x}_j &= \frac{\Sigma X_{cwc}}{n_j} = \frac{n_{0j}(-p_{1j}) + n_{1j}(p_{0j})}{n_j} \\ &= -p_{0j}p_{1j} + p_{1j}p_{0j} = 0 \quad \text{(B5)}\end{aligned}$$

Because $\bar{x}_j = 0$ after applying CWC, Equation 7 can be used to verify that β_{0j} is interpreted as the unadjusted (i.e., weighted) mean for cluster j.

When using a binary effect code, the mean for cluster j is $(n_{1j} - n_{0j})/n_j = p_{1j} - p_{0j}$, and X_{cwc} can be written as $x_{ij} - \bar{x}_j = x_{ij} - (p_{1j} - p_{0j})$. Simplifying this expression yields a centered score of $-2p_{1j}$ for the base group cases (i.e., $x_{ij} = -1$), and a score of $2p_{0j}$ for the comparison group cases (i.e., $x_{ij} = 1$). These centered values are identical to those from the dummy code scheme, except for the presence of a multiplicative constant. If the centered effect codes are inserted into Equation B5, the additional constant terms cancel, resulting in $\bar{x}_j = 0$. As such, β_{0j} is interpreted as the unadjusted mean for cluster j, regardless of whether one uses dummy or effect codes to represent the binary variable.

Received February 8, 2006
Revision received January 24, 2007
Accepted January 25, 2007 ■