The Application of Hierarchical Linear Modeling to Organizational Research

David A. Hofmann
Mark A. Griffin
Mark B. Gavin

Hierarchically ordered systems are an integral and defining aspect of organizations. How one chooses to investigate these hierarchically ordered systems has been discussed in a number of disciplines (sociology, economics, education, marketing, management, and psychology) for quite some time. The principal issue of concern is how to analyze and interpret data that reside at different levels of analysis. Hierarchical linear models provide a conceptual and statistical mechanism for investigating and drawing conclusions regarding relationships that cross levels of analysis. This chapter discusses the hierarchical nature of organizations and organizational research, provides an overview of the hierarchical linear modeling (HLM) methodology, and, using example research questions and simulated data, provides a step-by-step analytical introduction to these models. In addition, a number of other more specific issues are discussed, including statistical power and sample-size requirements, the implications of the statistical assumptions underlying these models, how the scaling of variables can change the interpretation and meaning of the estimated model,
and how hierarchical linear models relate to other multilevel analytic approaches.

The inherent hierarchical order and structure of organizations is apparent throughout the various theoretical and methodological contributions to this volume. Hall (1987) noted implicitly this hierarchical structure when he defined organizations as individuals organized into a collective residing in an environment—a definition that encompasses three hierarchical levels consisting of individuals, a collective or system, and an environment. For organizational researchers, this hierarchical structure manifests itself in the observation of hierarchical relationships. Hierarchical relationships occur when events at one level of analysis influence or are influenced by events at another level of analysis. Given the structure of organizations, a wide variety of hierarchical relationships can be observed and, in fact, have frequently been discussed and investigated in the organizational sciences. For example, researchers have investigated the relationships between organizational environmental factors and organizational structures (Aldrich & Pfeffer, 1976; Pfeffer & Salancik, 1978), organizational/subunit technologies and individual attitudes (Hulin & Roznowski, 1985), group norms/stimuli and individual behavior (Hackman, 1992), departmental characteristics/structure and individual attitudes (Brass, 1981; James & Jones, 1976; Oldham & Hackman, 1981; Rousseau, 1978), and climate/culture and individual behavior (Hofmann & Martins, 1992; James, James, & Aske, 1990; Kozlowski & Hults, 1987; Martocchio, 1994). In each of these examples, at least two different levels of analysis were involved.

Although it might not be initially obvious, hierarchical structures also occur in longitudinal studies when, for example, a time series of measurements is taken on a variable of interest for a number of different units (Nesselroade, 1991). Several researchers have suggested that such longitudinal data can be investigated by first analyzing patterns of change within units over time and then relating these patterns to between-unit variables (Bryk & Raudenbush, 1987; Deadrick, Bennett, & Russell, 1997; Eyring, Johnson, & Francis, 1993; Hofmann, Jacobs, & Gerras, 1992; Hofmann, Jacobs, & Baratta, 1993). Organizational studies of change that could be considered from this perspective range from micro investigations of individual work performance to macro investigations of organizational growth and decline (Austin & Villanova, 1992; Child, 1974, 1975; Deadrick et al., 1997; Hofmann et al., 1992, 1993). For example, at the organizational level, Child (1974, 1975) has investigated managerial influences on organizational performance, using a sample of eighty-two British firms spanning six industries. Although Child examined predictors of the average growth in income, assets, and sales over a five-year period by averaging these performance indices, an alternative approach would have been first to investigate the change pattern for each firm over time and then investigate predictors of these change patterns. This type of approach would have taken into account the nested and hierarchical nature of the data (that is, nested time series within firms).

As can be seen from the diversity of the examples just cited, hierarchical data structures occur in a large number of content areas within organizational science. As will be seen, the principal benefit of using multilevel methods is that they allow researchers to examine the relationship between variables that span different levels of analysis. Often in the past these hierarchical data structures have been either implicitly or explicitly ignored. In this chapter, we provide an introduction to and overview of hierarchical linear models—an approach to multilevel data that greatly facilitates the explicit recognition and investigation of these types of data structures.

Overview of Hierarchical Linear Modeling Methodology

Consistent with the hierarchical linear modeling approach, we will discuss an example in which a researcher is interested in investigating a dependent variable at the lowest level of analysis and independent variables at the same level of analysis as well as at higher levels. These types of research questions have been referred to as either cross-level (Rousseau, 1985) or mixed-determinant models (Klein, Dansereau, & Hall, 1994). It should be noted that the choice to focus on these models is more than coincidental. In fact, it is for these types of questions that hierarchical linear models are best suited. Although there has been some initial work using these models to investigate upward relationships (Griffin, 1997a), the traditional application will be to research questions taking the form of those described in this chapter.
Dealing with Hierarchical Data

As an example of this type of model, say that a researcher is interested in the relationships among individual employees' helping behavior toward their co-workers, their current affective state (that is, mood), and the level of work-group cohesion. We will assume, for this example, that work-group cohesion is a group-level variable that has been measured by having each group member complete a survey, and that sufficient within-group agreement and between-group variance exist to warrant the aggregation of these individual perceptions to the group level (Bliwise, Chapter Eight, this volume). Therefore, the focus will be on the group mean as an indicator of the work group's level of cohesion. In this case, individual-level helping behavior is the dependent variable, and mood and cohesion are the independent variables. Notice that, given our assumptions about the level of analysis of our variables, these independent variables span two levels of analysis; that is, mood is an individual-level variable, and cohesion is a group-level variable. When one is confronted with a research design and variables of this kind, there are primarily three options from which one can choose.

First, one can disaggregate the data so that the lower-level units are assigned a score representing their value on the lower-level variable. In our example, this would take the form of assigning to each individual a score representing the level of the work group's cohesion. In other words, everyone in the group would receive the same cohesion score. After this is done for each group included in the study, one can proceed with traditional ordinary least squares (OLS) regression analysis, investigating mood and cohesion as predictors of individual-level helping behavior. As we will discuss in our comparison between OLS regression and hierarchical linear models, this approach has several shortcomings, including the likely violation of statistical assumptions.

The second major option is to aggregate the lower-level variables to the same level as the higher-level variables. In our example, this would entail aggregating individual-level helping behavior as well as mood to the group level of analysis (that is, the level of analysis at which the cohesion measure resides) and then using the group means of these variables in subsequent analyses. The disadvantage of this approach, however, is that potentially meaningful individual-level variance is ignored both in the outcome measure and in one of the predictors. This ignored variance may indeed be meaningful and informative. In addition, if there is meaningful individual-level variance (in mood, for example), then the aggregation of this variable to represent a group-level property (for example, group affective tone) might result in a group-level variable with questionable construct validity (Klein et al., 1994).

Hierarchical linear models represent the third major option for dealing with hierarchically nested data structures. These models are specifically designed to overcome the weaknesses of the disaggregated and aggregated approaches discussed above. For one thing, they explicitly recognize that individuals within a group may be more similar to one another than they are to individuals in another group and may not, therefore, provide independent observations. In other words, these approaches explicitly model both the lower-level and the higher-level random-error components, therefore recognizing the partial interdependence of individuals within the same groups. This aspect of hierarchical linear models is in contrast to OLS approaches, where individual- and group-level random errors are not separately estimated. In addition, these models allow one to investigate both lower-level and higher-level variance in the outcome variable while maintaining the appropriate level of analysis for the independent variables.

A Brief Overview of Hierarchical Linear Models

The hierarchical linear modeling approach is a two-stage strategy that investigates variables occurring at two levels of analysis. (Software is available to analyze more than two levels of analysis. For the sake of clarity of presentation, however, we will focus on two-level models.) In the first stage, or level 1 analysis, relationships among level 1 variables are estimated separately for each higher-level unit. With respect to our example, this entails regressing individual-level helping behavior onto mood for each group. The outcome of this first stage is intercept and slope terms estimated separately for each group. Thus, for each group there will be a level 1 intercept term as well as a slope term summarizing the relationship between mood and helping behavior. These intercept and slope estimates
from the level 1 analysis are then used as outcome variables in the level 2 analysis. In our example, we will investigate the degree to which cohesion predicts the variance across groups in the level 1 intercepts and slopes.

Viewing hierarchical linear models in equation form, the level 1 component of our example regresses individual level helping behavior onto individual levels of mood separately for each group:

\[
\text{Helping behavior}_{ij} = \beta_{0j} + \beta_{ij} \cdot \text{Mood}_{ij} + r_{ij}
\]

where \(\text{Helping behavior}_{ij}\) is the degree of helping of individual \(i\) in group \(j\), \(\text{Mood}_{ij}\) is the mood score for the same individual, and \(r_{ij}\) represents random individual error. \(\beta_{0j}\) is the intercept value for group \(j\), and \(\beta_{ij}\) is the regression slope for group \(j\).

Equation 1 differs from the usual regression equation in that the parameters are estimated for each of the \(j\) groups separately. For this reason, the parameters can vary across groups. The subscript \(j\) associated with the parameters indicates that the parameter estimates can have a different value for each group.

In the second stage of this approach, the level 1 parameters are used as dependent variables for analysis at the group level. In equation format, and using our ongoing example, one can envision the following model:

\[
\beta_{ij} = \gamma_{00} + \gamma_{01} \cdot \text{Cohesion}_j + U_{0j}
\]

\[
\beta_{ij} = \gamma_{10} + \gamma_{11} \cdot \text{Cohesion}_j + U_{1j}
\]

where \(\beta_{ij}\) and \(\beta_{ij}\) are defined as before (that is, as the level 1 intercept and slope respectively), \(\text{Cohesion}_j\) is a group-level measure of the psychological attraction among group members, \(\gamma_{00}\) and \(\gamma_{10}\) are level 2 intercept terms, \(\gamma_{01}\) and \(\gamma_{11}\) are level 2 slope terms, and \(U_{0j}\) and \(U_{1j}\) are level 2 residuals. Equation 2 represents the main-effect model, which investigates whether cohesion is related to the between-group variance in helping behavior after controlling for mood. Equation 3 represents a cross-level interaction in such a way as to assess the degree to which cohesion moderates the within-group relationship between mood and helping behavior. Therefore, one could summarize this approach as a regression of regressions because the level 1 regression parameters (that is, intercepts and slopes) are themselves regressed onto higher-level variables in the level 2 analysis (Arnold, 1992).

Hierarchical Linear Models: History, Background, and Applications

Now that a brief overview of hierarchical linear models has been provided, it might be useful to discuss the literatures within which this approach has been applied. Even though the problems associated with hierarchically ordered data have been discussed in a number of different literatures—for example, in sociology (Blalock, 1984; Mason, Wong, & Entwistle, 1983), economics (Hanushek, 1974; Saxonhouse, 1976), biology (Laird & Ware, 1982), marketing (Wittink, 1977), and statistics (Longford, 1993)—the approach described here primarily emerged from educational research, where students are nested in classrooms and classrooms are nested in schools. In this case, researchers are primarily interested in assessing student as well as classroom and school effects on student learning and performance. Clearly, these data are multilevel in nature. In addition, the outcome variable is usually at the lowest level of analysis (that is, the student level), with both lower-level and higher-level predictors (that is, student, classroom, and school characteristics).

The relatively modern precursor of the present class of hierarchical linear models emerged from a conference convened in 1976 on data-aggregation problems in educational research (Burstein & Hannan, 1976; Burstein, Kim, & Delandshere, 1989). One of the outcomes of this conference was a proposal by Burstein (1980) to use a slopes-as-outcomes model; hence the initiation of the general form of the multilevel model presented earlier. Although statistical estimation problems initially stalled the full development of these models, Raudenbush (1988) noted a series of statistical developments that made these models more technically feasible by the middle of the 1980s (see also Bryk & Raudenbush, 1992). Under several different labels—hierarchical linear models (Bryk & Raudenbush, 1992), multilevel linear models (Goldstein, 1987), variance-components models (Longford, 1986), and random-coefficient models (Longford, 1993)—these models have become increasingly complex. As a result,
specialized software packages have been developed and are commonly available (Kreft, DeLeeuw, & van der Leeden, 1994). In the discussion that follows, we use hierarchical linear models as a broad term encompassing this general approach to multilevel data. When we use the HLM nomenclature, we will be specifically referring to the HLM software (Bryk, Raudenbush, & Congdon, 1996).

Since the widespread development of these models, they have been applied in a number of different research domains outside the organizational sciences. For example, recent applications and discussions of hierarchical linear models have spanned the following domains and topics:

- Meta-analysis (Kalaian & Raudenbush, 1996)
- Psychological change and distress in married couples (Barnett, Raudenbush, Brennan, Pleck, & Marshall, 1995; Raudenbush, Brennan, & Barnett, 1995)
- Psychotherapy (Joyce & Piper, 1996)
- Cognitive growth (Plewis, 1996)
- Information processing after a head injury (Zwaagstra, Schmidt, & Vanier, 1996)
- Educational research (Battistich, Solomon, Kim, Watson, & Schaps, 1995; Raudenbush, Rowan, & Cheong, 1993)
- Sociology (Jones, 1995)
- Program evaluation (Osgood & Smith, 1995)

In the organizational sciences, these models have been used to investigate questions arising in a number of different substantive domains. For example, they have been used to investigate the following topics:

- Goal congruence (Vancouver, Millsap, & Peters, 1994)
- The interaction of environment, person, and behavior (Vancouver, 1997)
- The interaction between individuals and situations (Griffin, 1997a)
- The dynamic nature of performance criteria (Deadrick et al., 1997; Hofmann et al., 1993)
- The moderating influence of leadership climate (Gavin & Hofmann, in press)
- Contextual effects on organizational citizenship behavior (Kidwell, Mossholder, & Bennett, 1997)
- Procedural-justice context (Mossholder, Bennett, & Martin, 1998)
- The effects of work group cohesiveness on performance and organizational commitment (Wech, Mossholder, Steel, & Bennett, 1998)
- Perceptions of client satisfaction with health services (Jimmieson & Griffin, 1998)
- Individual skill acquisition (Eyring et al., 1993; Gully, 1997)
- Individual and organizational predictors of pay satisfaction (Griffin, 1997b).

Although this list of topics does convey the growing popularity of these models, it might be useful to review one or two of these studies in more depth, to give an idea of the types of questions and interpretations that they allow. Kidwell and colleagues (1997), for example, report a study investigating the relationship between group cohesion and organizational citizenship behavior. Specifically, they proposed that citizenship behaviors within work groups would be influenced by individual differences in job satisfaction and by group-level differences in cohesion. They also proposed that group cohesion would be a cross-level moderator so that the relationship between job satisfaction and citizenship behavior would be stronger in groups that were more cohesive.

Using a sample of 260 employees in forty-nine work groups, they found that citizenship behavior varied systematically across groups, and that the level of group cohesion predicted differences in citizenship behavior even after individual differences in job satisfaction were controlled for. They also found that the relationship between satisfaction and citizenship behavior varied across groups, and that this relationship was stronger when groups were more cohesive. In other words, the relationship between job satisfaction and citizenship behavior was accentuated in more cohesive work groups. Thus work-group cohesion acted as a cross-level moderator of the relationship between job satisfaction and citizenship behavior.

Another example, from a longitudinal perspective, is a recent paper by Deadrick and colleagues (1997). These authors have investigated the dynamic nature of individual performance over time. When longitudinal data are investigated with hierarchical linear models, the level 1 model becomes a within-person analysis...
summarizing an individual's performance trajectory over time (see also Hofmann et al., 1993). The level 2 model then investigates individual-level predictors of these performance trajectories. In their study, Deadrick and colleagues (1997) investigated the performance of sewing-machine operators over the first twenty-four weeks of employment. They found significant differences across individuals, both in initial performance (that is, the performance-trajectory intercept) and in linear trend, or change in performance over time (that is, the linear slope of the performance trajectory). With respect to predicting individual differences on these trajectories, the authors found that cognitive ability and previous job-related experience predicted both initial performance and linear change in performance over time. Specifically, the results indicated that those sewing-machine operators with higher cognitive abilities and more previous job-related experience had higher initial job performance but improved less over time, which suggests that there may have been a ceiling effect with respect to performance over time.

As can be seen by the breadth of applications already noted, hierarchical linear models constitute a generalized tool for investigating multilevel relationships. Although this approach to multilevel data emerged primarily out of the education literature, it has more recently expanded and been integrated into a number of different literatures. Although its use in the organizational sciences has been increasing only recently, the widespread applicability and generalized nature of this approach, coupled with the nested nature of much organizational data, should make hierarchical linear models more popular in the future.

**Using Hierarchical Linear Models to Investigate Substantive Research Questions**

Now that a brief overview of hierarchical linear models has been presented, along with several substantive examples, it might be useful to consider in greater depth the process a researcher would likely follow in using this approach to investigate substantive questions. In this section, we return to our ongoing substantive example and specify a set of possible research questions and hypotheses. We then walk through the hierarchical linear modeling approach to testing these hypotheses. (This section is based, in part, on the overview provided by Hofmann, 1997.)

To illustrate the procedure for testing the hypotheses using the HLM program, we generated simulated data consisting of fifty groups with twenty individuals in each one. (The data used to illustrate the HLM procedure are available from any of the three coauthors.) Recall that our substantive example is focused on the relationship among helping behavior, mood, and work-group cohesion. Furthermore, helping behavior and mood are assumed to be individual-level variables, whereas work-group cohesion is assumed to be a group-level variable. The content of these hypotheses is as follows:

H1. **Individual-level mood will be positively related to helping behavior.**

H2. **Cohesion will be positively related to helping behavior after controlling for individual-level mood.**

H3. **Cohesion will moderate the relationship between mood and helping behavior so that mood and helping behavior will be more strongly related when the group is more cohesive.**

In the following section, we will describe how to test these hypotheses with the HLM program.

**Hierarchical Linear Models: Equations, Effects, and Statistical Tests**

Before detailing the sequence of models estimated to investigate these hypotheses, it is necessary to define some of the basic components of the hierarchical linear model. In our brief overview of hierarchical linear models, we presented three equations (see equations 1, 2, and 3).

Although the level 1 and level 2 models were discussed earlier as separate equations, it should be noted that they are estimated simultaneously. Three key terms that arise in estimating these models are *fixed effects*, *random coefficients*, and *variance components*. Fixed effects are parameter estimates that do not vary across groups. For example, equations 2 and 3 are estimated at the group level (that is, across groups) and so there is only one parameter that summarizes the relationship across all groups. As a result, the parameters estimated in these equations do not vary across groups. The parameters in equations 2 and 3 are therefore fixed effects ($\gamma_{00}$, $\gamma_{01}$, $\gamma_{10}$, and $\gamma_{11}$).
Hierarchical linear models, and the HLM software specifically (Bryk et al., 1996), estimate these fixed effects by using a generalized least squares (GLS) regression approach. Although these level 2 regression parameters could be estimated with an OLS approach, this is not appropriate, given that the precision of the level 1 parameters (that is, the level 2 dependent variable) will likely vary across groups. In other words, because the standard errors of the level 1 parameters can vary across groups, the reliability of the level 2 outcome variable can be different for each group. Some level 1 parameters will be better estimates of the underlying relationship, and it is this variation in precision that is taken into account in the level 2 analysis. This GLS estimation provides a weighted level 2 regression so that groups with more reliable (that is, more precise) level 1 estimates receive more weight and therefore have more influence in the level 2 regression; t-test statistical tests are provided for these fixed effects.

The variance of the level 1 residuals (that is, the variance in r_{ij}) and the variance-covariance of the level 2 residuals (that is, the variance-covariance matrix of U_{0j} and U_{1j}) comprise the variance components. The variances and covariances of the level 2 residuals are contained in the \( \tau \) matrix. The HLM procedure uses the EM algorithm (Dempster, Laird, & Rubin, 1977; Raudenbush, 1988) to produce maximum-likelihood estimates of the variance components. With regard to statistical tests, HLM provides a chi-square test for the level 2 residual variances—in our example, \( U_{0j} \) and \( U_{1j} \)—assessing whether the particular variance component departs significantly from zero.

Random coefficients are those coefficients that are allowed to vary across groups. In the current example, the level 1 intercepts and slopes (that is, \( \beta_{0j} \) and \( \beta_{1j} \)) are random coefficients. The HLM procedure does not provide any statistical tests for these parameters. However, as will be shown, one can assess whether the mean and variance of these parameters depart significantly from zero.

One-Way Analysis of Variance: Partitioning Outcome Variance into “Within” and “Between” Components

In looking at the three example hypotheses listed earlier, one thing to notice is that helping behavior is hypothesized to be predicted by both individual- and group-level independent variables. For these hypotheses to be supported, there must be variation in helping behavior at both the individual and the group level. In other words, if group-level variables are going to be significantly related to helping behavior, then helping behavior, by definition, must contain meaningful between-group variance. This first model investigates the amount of between-group variance in helping behavior by partitioning the total variance in helping behavior into its within-group and between-group components. This model is conceptually equivalent to a one-way analysis of variance (ANOVA), with helping behavior as the dependent variable and with group membership serving as the independent variable. To do this, the following equations can be estimated:

Level 1: \[ \text{Helping}_{ij} + \beta_{0j} + \epsilon_{ij} \]
Level 2: \[ \beta_{0j} = \gamma_{00} + U_{0j} \]

where

\( \beta_{0j} = \text{mean for helping for group } j \)
\( \gamma_{00} = \text{grand mean helping} \)
Variance (\( r_{ij} \)) = \( \sigma^2 \) = within-group variance in helping
Variance (\( U_{0j} \)) = \( \tau_{00} \) = between-group variance in helping

In this set of equations, the level 1 equation includes no predictors; therefore, the regression equation includes only an intercept term. To compute intercept terms in regression, the analysis includes a unit vector as a predictor in the equation. The parameter associated with this unit vector represents the intercept term in the final regression equation. In regression software packages, the researcher typically does not explicitly model this unit vector. Similarly for hierarchical linear models, when a researcher specifies no predictors in a level 1 or level 2 equation, the variance in the outcome measure is implicitly regressed onto a unit vector, producing a regression-based intercept estimate. In the level 1 equation, the \( \beta_{0j} \) parameter will be equal to that group’s mean level of helping behavior (that is, if a variable is regressed only onto a constant unit vector, the resulting parameter is equal to the mean).
The level 2 model regresses each group's mean helping behavior onto a constant; that is, $\beta_0j$ is regressed onto a unit vector, resulting in a $\gamma_{00}$ parameter equal to the grand mean helping behavior (the mean of the group means, $\beta_0j$). Given that each of the respective dependent variables is regressed onto a constant, it follows that any within-group variance in helping behavior is forced into the level 1 residual ($r_{ij}$) and any between-group variance in helping behavior is forced into the level 2 residual ($U_{0j}$).

Although HLM does not provide a significance test for the within-group variance component ($\sigma^2$), it does provide a significance test for the between-group variance ($\tau_{00}$). In addition, the ratio of the between-group variance to the total variance can be described as an intraclass correlation (Bryk & Raudenbush, 1992). In the preceding model, the total variance in helping behavior has been decomposed into its within- and between-group components:

\[
\text{Variance (Helping}_{ij} = \text{Variance (U}_{0j} + r_{ij} = \tau_{00} + \sigma^2
\]

Therefore, an intraclass correlation can be computed by investigating the following ratio:

\[
\text{ICC} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}
\]

This intraclass correlation represents a ratio of the between-group variance in helping behavior to the total variance in helping behavior. In summary, the one-way ANOVA provides the following pieces of information regarding the helping behavior measure:

1. The amount of variance residing within groups
2. The amount and significance of variance residing between groups
3. The intraclass correlation specifying the percentage of the total variance residing between groups (see Blesse, Chapter Eight, this volume, for a more detailed discussion of intraclass correlations)

Table 11.1 presents the results of this analysis, using the example data. The grand mean of helping behavior—that is, the average helping behavior pooled across individuals and groups—is 31.39:

\[
t(49) = 22.05, p < .01
\]

The within-group variance on the helping-behavior measure is 31.76 ($\sigma^2$), whereas the between-group variance in helping is 99.82.

The chi-square test indicates that this between-group variance is significant:

\[
\chi^2(49) = 3,129.25, p < .01
\]

The intraclass correlation for the helping-behavior measure is .76—that is, 99.82 / (99.82 + 31.76)—indicating that 76 percent of the variance resides between groups.

### Random-Coefficient Regression Model

The presence of significant between-group variance in helping behavior is the first requirement for testing hypotheses 2 and 3. In addition, these hypotheses propose that group-level cohesion will be significantly associated with the variance in the level 1 intercepts (hypothesis 2) and variance in the slopes (hypothesis 3). For these hypotheses to be supported, there must be significant variance in intercepts and slopes across groups. In other words, significant variance in the intercepts and slopes can be considered a precondition for the testing of hypotheses 2 and 3. The following model is designed to test these preconditions. In addition, this model will also directly test hypothesis 1. This model is called the random-coefficient regression model because it is similar to ordinary regression analysis, but in this model the coefficients are allowed to vary across groups. It takes on the following form:

**Level 1:**

\[
\text{Helping}_{ij} = \beta_0j + \beta_{1j} (\text{Mood}_{ij}) + r_{ij}
\]

**Level 2:**

\[
\begin{align*}
\beta_0j &= \gamma_{00} + U_{0j} \\
\beta_{1j} &= \gamma_{10} + U_{1j}
\end{align*}
\]

where

\[
\begin{align*}
\gamma_{00} &= \text{mean of the intercepts across groups} \\
\gamma_{10} &= \text{mean of the slopes across groups (hypothesis 1)} \\
\text{variance (r}_{ij} &= \sigma^2 = \text{level 1 residual variance} \\
\text{Variance (U}_{0j} &= \tau_{00} = \text{variance in intercepts} \\
\text{Variance (U}_{1j} &= \tau_{11} = \text{variance in slopes}
\end{align*}
\]
Table 11.1. Results of the Estimated Models, Based on Example Data Set.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\gamma_{00}$</th>
<th>$\gamma_{01}$</th>
<th>$\gamma_{10}$</th>
<th>$\gamma_{11}$</th>
<th>$\sigma^2$</th>
<th>$\tau_{00}$</th>
<th>$\tau_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-way ANOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1: Helping $\gamma_i = \beta_{0i} + r_{ij}$</td>
<td>31.39</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>31.76</td>
<td>99.82</td>
<td>—</td>
</tr>
<tr>
<td>L2: $\beta_{0j} = \gamma_{00} + U_{0j}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random-coefficient regression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1: Helping $\gamma_i = \beta_{0i} + \beta_{ij} \cdot \text{Mood}<em>{ij} + r</em>{ij}$</td>
<td>31.42</td>
<td>—</td>
<td>3.01</td>
<td>—</td>
<td>5.61</td>
<td>45.63</td>
<td>.13</td>
</tr>
<tr>
<td>L2: $\beta_{0j} = \gamma_{10} + U_{ij}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercepts-as-outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1: Helping $\gamma_i = \beta_{0i} + \beta_{ij} \cdot \text{Mood}<em>{ij} + r</em>{ij}$</td>
<td>24.92</td>
<td>1.24</td>
<td>3.01</td>
<td>—</td>
<td>5.61</td>
<td>41.68</td>
<td>.13</td>
</tr>
<tr>
<td>L2: $\beta_{0j} = \gamma_{10} + U_{ij}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slopes-as-outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1: Helping $\gamma_i = \beta_{0i} + \beta_{ij} \cdot \text{Mood}<em>{ij} + r</em>{ij}$</td>
<td>25.14</td>
<td>1.19</td>
<td>2.06</td>
<td>.18</td>
<td>5.61</td>
<td>42.95</td>
<td>.02</td>
</tr>
<tr>
<td>L2: $\beta_{ij} = \gamma_{10} + \gamma_{11} \cdot \text{Cohesion}<em>{ij} + U</em>{ij}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Parameters defined as follows:

$\gamma_{00}$ = Intercept of level-2 regression predicting $\beta_{0i}$

$\gamma_{01}$ = Slope of level-2 regression predicting $\beta_{0j}$

$\gamma_{10}$ = Intercept of level-2 regression predicting $\beta_{ij}$

$\gamma_{11}$ = Slope of level-2 regression predicting $\beta_{ij}$

$\sigma^2$ = Variance in level-1 residual (i.e., variance in $r_{ij}$)

$\tau_{00}$ = Variance in level-2 residual for models predicting $\beta_{0j}$ (i.e., variance in $U_{0j}$)

$\tau_{11}$ = Variance in level-2 residual for models predicting $\beta_{ij}$ (i.e., variance in $U_{ij}$)

The direction of the pooled slope of the regression of helping behavior onto mood ($\gamma_{00}$) indicates that, on average, individuals in a more positive mood are more likely to help their coworkers.

$\gamma_{00} = 31.42, t(49) = 32.74, p < .01$

The variance components differ significantly from zero, and in this case, the random-coefficient regression model provides a significant test for the variance in the level-1 regression coefficients as well as for the mean of the level-1 regression coefficients. Table 11.1 provides the results for the random-coefficient regression model for the example data. Both $\gamma_{00}$ and $\gamma_{10}$ indicate the average of the pooled intercept slope is significantly different from zero.

$\gamma_{10} = 3.01, t(49) = 43.65, p < .01$

Because there are no level-2 predictors of either $\beta_{0j}$ or $\beta_{ij}$, the level 2 regression equation is simply equal to an intercept term and a residual. In this form, the $\gamma_{00}$ and $\gamma_{01}$ parameters represented the variance of the intercepts ($\tau_{00}$ and $\tau_{10}$ parameters). Similarly, given that $\beta_{0j}$ and $\beta_{ij}$ are residuals, the variance of the slopes ($\tau_{11}$ parameter) is also equal to the variance of the slopes ($\tau_{10}$ and $\tau_{11}$ parameters). HLM provides a test related to the variance in the level 2 residual terms ($\tau_{00}$ and $\tau_{11}$ parameters). This test provides a direct test of zero. In other words, the test determines whether the overall variance across groups is significant. In this case, the level 2 residual term $\tau_{00}$ is significantly different from zero. The variance of the slopes ($\tau_{10}$ parameter) is also equal to the variance of the slopes ($\tau_{10}$ and $\tau_{11}$ parameters). HLM also provides a chi-square test for the two residual variances ($\tau_{00}$ and $\tau_{11}$). These chi-square tests indicate whether the variances ($\tau_{00}$ and $\tau_{11}$) are significant.
thereby supporting hypothesis 1. The results in Table 11.1 also illustrate that, across groups, there is significant variance in the intercepts:

\[ \tau_{00} = 45.63, \chi^2(49) = 4,605.68, p < .01 \]

There is also significant variance, across groups, in the slopes:

\[ \tau_{11} = .13, \chi^2(49) = 110.17, p < .01 \]

Thus the preconditions for hypotheses 2 and 3 are also supported.

The random-coefficient regression model allows us to do one more thing. Although we know that, on average, mood is significantly related to helping behavior, we do not know the magnitude of this relationship. However, we do have from the ANOVA model an estimate of the within-group variance in helping behavior (\(\sigma^2 = 31.76\)), and from the current model we have an estimate of the residual within-group variance after controlling for mood (\(\sigma^2 = 5.61\)). Comparing these two variance estimates allows us to compute an \(R^2\) for the relationship between helping behavior and mood. Specifically, one can obtain the \(R^2\) for helping behavior by computing the following ratio:

\[ R^2 \text{ for level 1 model} = \frac{(\sigma^2\text{ oneway ANOVA} - \sigma^2\text{random regression})}{\sigma^2\text{oneway ANOVA}} \]

This ratio compares the amount of variance accounted for by mood to the total within-group variance in helping in the denominator. Therefore, this ratio represents the percentage of the level 1 variance in helping that is accounted for by mood. In our example data, this \(R^2\) is equal to .82:

\[ \frac{(31.76 - 5.61)}{31.76} \]

It should be emphasized that this \(R^2\) value is computed with the use of the within-group variance as the denominator, not with the use of the total variance in the outcome variable (see Snijders & Bosker, 1994, for several alternative computations of \(R^2\) values). Thus it is important to keep in mind that this \(R^2\) value, as well as those to be discussed, are computed relative to the variance that can be predicted by a given independent variable; they are not computed relative to the total variance in the outcome variable. For example, level 1 variables are evaluated versus the within-group variance in the outcome variable, whereas level 2 variables (as will be shown) are evaluated relative to the between-group variance in the intercepts and slopes, respectively.

### Intercepts-as-Outcomes Model

Given that our example data demonstrated significant between-group variance in the intercept term across groups, the next step is to see if this variance is significantly related to work-group cohesion (hypothesis 2). The HLM model takes the following form:

**Level 1:** \[ \text{Helping}_{ij} = \beta_{0j} + \beta_{ij}(\text{Mood}_{ij}) + r_{ij} \]

**Level 2:** \[ \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Cohesion}_{j}) + U_{0j} \]

\[ \beta_{ij} = \gamma_{10} + U_{ij} \]

where

- \(\gamma_{00}\) = level 2 intercept
- \(\gamma_{01}\) = level 2 slope (hypothesis 2)
- \(\gamma_{10}\) = mean (pooled) slopes
- Variance \((r_{ij}) = \sigma^2 = \text{level 1 residual variance}\)
- Variance \((U_{0j}) = \tau_{00} = \text{residual intercept variance}\)
- Variance \((U_{ij}) = \tau_{11} = \text{variance in slopes}\)

This model is similar to the random-coefficient regression model discussed earlier, with the addition of cohesion as level 2 predictor of \(\beta_{0j}\). The \(t\)-tests associated with the \(\gamma_{01}\) parameter provide a direct test of hypothesis 2. Given that the level 2 equation for \(\beta_{0j}\) now includes a predictor (cohesion), the variance in the \(U_{0j}\) parameter \((\tau_{00})\) represents the residual variance in \(\beta_{0j}\) across groups. If the chi-square test for this parameter is significant, it indicates that there remains systematic level 2 variance that could be modeled by additional level 2 predictors. All other parameters take on the same
meaning as they did under the estimation of the random-coefficient regression model.

Inspection of Table 11.1 indicates that, in our example data, cohesion is a significant predictor of the between-group variance in the intercept term:

\[ \gamma_{10} = 1.24; t(48) = 2.47, p = .02 \]

The chi-square test associated with the residual variance in the intercept across groups indicates that there is still significant variance remaining in this parameter across groups:

\[ \tau_{00} = 41.68; \chi^2(48) = 4,127.70, p < .01 \]

The \( \gamma_{10} \) parameter takes on the same meaning (and same value) as in the previous model and indicates that, on average, mood is significantly and positively related to helping behavior. The chisquare test for the variance in the \( \tau_{11} \) indicates that there is significant variance in the slope term:

\[ \tau_{11} = .13, \chi^2(49) = 110.18, p < .01 \]

The modeling of this variance and the testing of hypothesis 3 is taken up in the next model.

After the estimation of the intercepts-as-outcomes model, we know that cohesion was significantly related to the variance in the intercept term across groups. In order to assess the magnitude of this relationship, we can compare the variance in the \( \tau_{00} \) from the random-coefficient regression model (the total between-group variance in the intercept term across groups) with the variance in the \( \tau_{00} \) for the current model (the residual variance in the intercept after accounting for cohesion). Specifically, by comparing these two \( \tau \)'s one can obtain the \( R^2 \) for cohesion by computing the following ratio:

\[
R^2 \text{ for level 2 intercept model} = \frac{(\tau_{00} \text{ random regression} - \tau_{00} \text{ intercepts-as-outcomes})}{\tau_{00} \text{ random regression}}
\]

For our example data, this \( R^2 \) is equal to .09—that is, \((45.63 - 41.68)/45.63\). Once again, note that this \( R^2 \) value is computed relative to the between-group variance in the intercepts; it is not computed relative to the total variance in the outcome variable (Snijders & Bosker, 1994).

Slopes-as-Outcomes Model

Given that the preceding model indicates significant variance in the level 1 slopes across groups (\( \tau_{11} \)), we can see if the variance is significantly related to cohesion (hypothesis 3). The HLM model takes the following form:

Level 1: \[ \text{Helping}_{ij} = \beta_{0j} + \beta_{1j} \text{ (Mood}_{ij} + r_{ij} \]

Level 2: \[ \beta_{0j} = \gamma_{00} + \gamma_{01} \text{ (Cohesion}_{j} + U_{0j} \]
\[ \beta_{1j} = \gamma_{10} + \gamma_{11} \text{ (Cohesion}_{j} + U_{1j} \]

where

\[ \gamma_{00} = \text{level 2 intercept} \]
\[ \gamma_{01} = \text{level 2 slope (hypothesis 2)} \]
\[ \gamma_{10} = \text{level 2 intercept} \]
\[ \gamma_{11} = \text{level 2 slope (hypothesis 3)} \]

Variance = \( r_{ij}^2 = \sigma^2 \) = level 1 residual variance

Variance = \( (U_{0j})^2 = \tau_{00} \) = residual intercept variance

Variance = \( (U_{1j})^2 = \tau_{11} \) = residual slope variance

The differences between this slopes-as-outcomes model and the intercepts-as-outcomes model are that cohesion is now included as a predictor of the \( \beta_{1j} \) parameter and, as a result, the \( U_{1j} \) variance is now the residual variance in the \( \beta_{1j} \) parameter across groups instead of the total variance across groups. Once again, if the chisquare test associated with this parameter variance is significant, it indicates that there remains systematic variance in the \( \beta_{1j} \) parameter that could be modeled by additional level 2 predictors. In
addition, the $t$-test associated with the $\gamma_{11}$ parameter provides a direct test of hypothesis 3. This hypothesis represents a cross-level moderator or cross-level interaction because a group-level variable is hypothesized to moderate the relationship between two individual-level variables.

Inspection of Table 11.1 indicates that, for our example data, $\gamma_{11}$ is significant, thereby supporting hypothesis 3:

$$\gamma_{11} = .18, t(48) = 6.32, p<.01$$

The positive parameter estimate indicates that as cohesion increases, the slope relating mood to helping behavior becomes more positive (that is, stronger). Table 11.1 also reveals that the remaining variance in the $\beta_{1j}$ parameter is not significantly different from zero:

$$\gamma_{11} = .02, \chi^2(48) = 59.22, ns$$

Once again, we use the value of the $\tau_{11}$ parameter from the previous model with the value of the $\tau_{11}$ parameter from the current model to compute an $R^2$ for cohesion as a level 2 moderator of the relationship between individual-level mood and helping behavior. Specifically, one can obtain the $R^2$ as follows:

$$R^2 \text{ level 2 slope model} = (\tau_{11} \text{ intercept-as-outcomes} - \tau_{11} \text{ slopes as outcomes}) / \tau_{11} \text{ intercept as outcomes}$$

For the example data, this $R^2$ value is .85—that is, $(.13 - .02) / .13$—which indicates that cohesion accounts for 85% of the variance in the relationship between mood and helping behavior.

Looking across the preceding models, the data have supported each of our three hypotheses, listed earlier. Specifically, the random-coefficient regression model provides support for hypothesis 1, the intercepts-as-outcomes model provides support for hypothesis 2, and the slopes-as-outcomes model provides support for hypothesis 3. Although the preceding sequence of models provides a test of a series of relatively simple hypotheses, the extension of these models to include more level 1 and level 2 predictors is relatively straightforward. Additional details regarding the testing of more complex models can be found in Bryk and Raudenbush (1992), Goldstein (1995), and Longford (1993).

**Key Assumptions of Hierarchical Linear Models**

Hierarchical linear models carry with them certain assumptions regarding the organizational systems under investigation, the data structures that can be analyzed, and the distributional properties of variables. In the passages that follow, we review some of the methodological and statistical assumptions on which these models are based.

**Methodological Assumptions**

The first assumption is that lower-level units (for example, individuals) are nested within identifiable higher-level units (for example, groups/teams or departments). In collecting data from such organizational systems, it is necessary that each lower-level unit be linked to an identifiable higher-level unit. A second assumption for both the organizational system and the data, the lower-level units are exposed to and influenced by characteristics and/or processes of the higher-level units. For example, in studying individuals nested within teams, it is assumed that team membership matters, and that individuals are influenced by characteristics and processes of the team. This is a general assumption underlying the hierarchical linear modeling approach. Whether a given characteristic and/or process of the higher-level unit has an impact on the lower-level outcome is, of course, an empirical question, which is tested via the estimation of the specific model(s).

A third assumption regarding the data is that the outcome variable is measured at the lowest level of interest to the researcher. In our example, for instance, the lowest level of interest is the individual level, and the outcome variable is individual-level helping behavior. Although the outcome variable will reside at the lowest level of interest to the researcher, the predictor variables can include both this level and higher levels of analysis. This is also illustrated in our example by the inclusion of both individual-level mood and group-level cohesion.
Finally, hierarchical linear models assume that the outcome variable varies both within the lower-level units and between the higher-level units. This is because these models investigate the influence of higher-level variables on lower-level outcomes, and this necessitates variance in the outcome both within and between units.

**Statistical Assumptions**

Hierarchical linear models, like any other statistical technique, require certain assumptions about the nature of the data. For two-level models, the following assumptions apply (Bryk & Raudenbush, 1992, p. 200):

1. Level 1 residuals are independent and normally distributed with a mean of zero and variance $\sigma^2$ for every level 1 unit within each level 2 unit.
2. Level 1 predictors are independent of level 1 residuals.
3. Random errors at level 2 are multivariate normal, each with a mean of zero, a variance of $\tau_{q0}$, and a covariance of $\tau_{q1}$, and are independent among level 2 units.
4. The set of level 2 predictors is independent of every level 2 residual. (This assumption is similar to assumption 2, but for level 2.)
5. Residuals at level 1 and level 2 are also independent.

To continue with our example, these assumptions entail the following meanings:

1. After taking into account the effect of mood, the within-group errors are normal and independent, with a mean of zero in each group and equal variances across groups (assumption 1).
2. If any additional level 1 predictors of helping are excluded from the model (and if their variance is thereby forced into the level 1 residual), they are independent of individual mood (assumption 2).
3. The group effects (that is, the level 2 residuals) are assumed multivariate normal, with variances $\tau_{00}$ and $\tau_{11}$ and covariance $\tau_{01}$ (assumption 3).

4. The effects of any group-level predictors excluded from the model for the intercept and mood slope are independent of cohesion (assumption 4).
5. The level 1 residual $r_{ij}$ is independent of the residual group effects $U_{0j}$ and $U_{1j}$.

Although Bryk and Raudenbush (1992) discuss these assumptions and the influence of possible violations, James (1995) noted several issues not discussed by Bryk and Raudenbush. First, hierarchical linear models assume multivariate normality and, on the basis of this assumption, proceed with maximum-likelihood estimation. The assumption of multivariate normality can be problematic, however, especially in the presence of interactions. This is clearly the case when level 1 slopes are predicted with level 2 variables. Second, hierarchical linear models treat independent variables as random variables; that is, processes beyond the control of the researcher determine the level of an individual's value on the independent variable. (This is in contrast with fixed variables, where individuals are randomly assigned to particular levels of the independent variable.) Given this assumption, it is possible that the independent variables will be correlated with the associated residuals. This could occur if an omitted variable is both correlated with the predictor variable included in the model and with the dependent variable (James, 1980). Finally, with regard to longitudinal data (Bryk & Raudenbush, 1987), hierarchical linear models assume that the level 1 residuals are independent, which may not be the case when one is modeling time-series data. Given the relative newness of hierarchical linear models, it remains to be seen how robust these techniques are to violations of these assumptions and, therefore, how robust this approach is to multilevel analysis.

**Additional Issues in the Use of Hierarchical Linear Models**

There are several additional issues that should be considered by researchers investigating multilevel relationships. In particular, the following section discusses the implications of alternative scalings of the level 1 predictors, as well as sample-size requirements.
Alternative Scales for Level 1 Predictors

Because hierarchical linear models use the level 1 parameters (that is, intercepts and slopes) as dependent variables in the level 2 equations, it is imperative that researchers understand the interpretation of these parameters. Regression textbooks (e.g., Cohen & Cohen, 1983) note that the slope parameter represents the expected increase in the outcome variable, given a unit increase in the independent variable, whereas the intercept represents the expected value of the outcome measure when all the independent variables take on a value of zero. In the organizational sciences, it is often the case that variables do not have meaningful, or true, zero points. For example, it means little for a person to have zero mood, job satisfaction, or organizational commitment, or for an organization to have zero structure, technology, centralization, or formalization. Therefore, it is useful to ask whether there are alternative scalings that would render the intercept more interpretable.

In hierarchical linear models, three alternative scalings of the level 1 independent variables have traditionally been discussed:

1. Raw-metric approaches, where the level 1 predictors are used in their original form
2. Grand-mean centering, where the grand mean of the level 1 variable is subtracted from each individual's score (for example, mood\(_{j} - \text{mood}_{\text{grand mean}}\))
3. Group-mean centering, where the group mean is subtracted from each individual's score on the predictor (for example, mood\(_{j} - \text{mood}_{\text{group mean}}\))

Although grand-mean and raw-metric approaches yield equivalent models, group-mean centering generally is not equivalent to these other two scaling options (Kreft, DeLeeuw, & Aiken, 1995). Recently, a number of researchers have discussed how group-mean centering versus raw-metric/ grand-mean centering options can change the estimation and interpretation of hierarchical linear models (Bryk & Raudenbush, 1992; Hofmann & Gavin, 1998; Kreft et al., 1995; Longford, 1989; Pwais, 1989; Raudenbush, 1989a, 1989b). In particular, these scaling options can influence the interpretation of level 2 intercept and slope models. Each of these will be discussed in turn.

Scaling Options and Level 2 Intercept Models

In level 2 intercept models, group-mean centering versus raw-metric/ grand-mean centering influences the interpretation of the variance of the level 2 intercept term. Specifically, when either grand-mean or raw-metric approaches are adopted, the variance in the intercept term represents the adjusted between-group variance in the outcome measure, after controlling for the level 1 predictors. In our example, and with these centering options, the variance in the intercept term across groups represents the between-group variance in helping behavior after controlling for mood. Given the wording of our three hypotheses, grand-mean or raw-metric centering of the level 1 predictors would be appropriate.

In using group-mean centering, alternatively, the level 1 intercept variance simply reflects the between-group variance in the outcome variable; that is, it reflects the unadjusted between-group variance in the outcome variable. In our example, and with group-mean centering, the intercept would reflect the between-group variance in helping behavior (the effects of individual mood have not been controlled for). Therefore, it would have been inappropriate for us to have tested hypothesis 2 by using group-mean centering, because all the effects of mood (within and between groups) would not have been taken into account. Put more simply, the level 2 model regressing the intercept term onto cohesion would be providing a test of the group-level relationship between helping and cohesion (equivalent to the group-level regression of the mean of helping behavior onto cohesion). Clearly, given the wording of hypothesis 2, this would have been an inappropriate analysis because the variance in mood—in particular, the between-group variance—has not been controlled for in the regression equation.

It is critical for researchers to understand that when group-mean centering is chosen as the scaling option for level 1 predictors, the between-group variance in these variables is excluded from the model. Therefore, the only way to completely control for the effects of level 1 variables (both within-group and between-group variance) under group-mean centering is to add this eliminated variance back in to the model. With respect to our example, the only way to control for the effects of mood before investigating the relationship between helping and cohesion would be to estimate a level 2 model with two predictors of helping behavior—namely, group-level mood and cohesion. This model would include the
between-group variance in mood that was eliminated via the group-mean centering and would therefore provide results similar to those of a grand-mean or raw-metric model. The important point to emphasize here is that when group-mean centering is used, the effects of the level 1 variables are not controlled for in estimating the level 2 models.

**Scaling Options and Level 2 Slope Models**

The distinction between group-mean centering and raw-metric/ grand-mean centering can also be important in level 2 slope models. This is true because, with raw-metric or grand-mean centering, the level 1 slope summarizing the relationship between the level 1 predictor and the outcome variable is actually a function of two different relationships: the within-group relationship between the predictor and the outcome, and the between-group relationship between the predictor and the outcome. To illustrate, let us assume that a researcher has an independent and a dependent variable, and that both contain meaningful variance within and between groups. Given this meaningful within- and between-group variance, the researcher could actually create three independent variables: one consisting of the within-group variance (group-mean–centered), one consisting of the between-group variance (group means), and one consisting of the raw metric, which would contain both within- and between-group variance. Three regression equations could then be estimated with the use of the within-group variance, the between-group variance, and the raw-metric variance. The first equation would provide an estimate of the within-group relationship, the second an estimate of the between-group relationship, and the third an estimate of the total, or composite, relationship.

This logic applies to hierarchical linear modeling as well. In particular, raw-metric–centered models or grand-mean–centered models produce level 1 slopes that are actually composites of the within-group and between-group relationships among the independent and dependent variable. In fact, it is only when one group mean centers the level 1 predictors (that is, eliminates the between-group variance in the level 1 predictor from the level 1 model) that the level 1 slopes provide a pure estimate of the within-group relationship (Raudenbush, 1989b). As Raudenbush (1989b) has pointed out, this inclusion of both the within-group relationship and the between-group relationship in the level 1 slope estimates, under either raw-metric or grand-mean centering, can result in spurious cross-level interactions (see Hofmann & Gavin, 1998, for an example using simulated data).

We think that researchers should at least investigate this possibility when evaluating cross-level interactions. For instance, in the case of our substantive example, we hypothesized that work-group cohesion would act as both a cross-level main effect and as a moderator—that is, it would predict both the level 2 intercepts and the slopes. We also hypothesized that the main effect of cohesion would incrementally predict helping behavior (that is, over and above the effect of mood). Thus raw-metric or grand-mean centering would be the appropriate choice for our main-effect model. But we also hypothesized that cohesion would act as a cross-level moderator as well, by significantly predicting the level 1 slopes. As already noted, raw-metric or grand-mean centering can, on occasion, produce spurious cross-level interactions (Hofmann & Gavin, 1998; Raudenbush, 1989b). With respect to practical recommendations, we would recommend that the researcher first estimate all the models (both level 2 intercept and slope models) by using either raw-metric or grand-mean centering and then estimate one additional model before concluding the analyses. This last model would specify group-mean centering for the level 1 predictor, add in the group mean of the level 1 predictor in the level 2 intercept model (to reintroduce the between-group variance in the level 1 predictor, variance that was eliminated via group-mean centering), and re-estimate the cross-level interaction. In the case of our substantive example, this final model would take on the following form:

Level 1: \( \text{Helping}_{ij} = \beta_{0j} + \beta_{1j} \text{Mood}_{i,j} \text{group centered} + r_{ij} \)

Level 2: \( \beta_{0j} = \gamma_{00} + \gamma_{01} \text{Mean Mood}_{i,j} + \gamma_{02} \text{Cohesion}_{i,j} + U_{0j} \)
\( \beta_{1j} = \gamma_{10} + \gamma_{11} \text{Cohesion}_{i,j} + U_{1j} \)
\( \beta_{2j} = \gamma_{20} + U_{2j} \)

If the \( \gamma_{ij} \) parameters across these two final models are virtually identical, then, in order to be consistent with the investigation of all other hypotheses, we would report the grand-mean–centered results and footnote that we checked the cross-level interaction, using group-mean centering to ensure that the result was not spurious. This way, the results are easier for the reader to follow because the
choice of centering is not changing halfway through the results section, but the researcher is confident that the results are not spurious.

It is important to emphasize that neither of these scaling options is statistically more correct than the other; instead, these decisions must be based on the theoretical model under consideration and on the nature of the hypotheses under investigation (for more details, see Hofmann & Gavin, 1998; Kreft et al., 1995). With respect to the organizational sciences, Hofmann and Gavin (1998) treat, at greater depth, these different centering options for different cross-level paradigms that traditionally have been investigated by organizational researchers.

**Sample-Size Requirements**

In the literature on hierarchical linear models, Kreft (1996) is, to our knowledge, the most comprehensive and complete summary of the work conducted on power and sample-size requirements. Summarizing simulation studies by Kim (1990) and Bassiri (1988), Kreft (1996) concludes that, in general, relatively large sample sizes are required. With regard to specific recommendations, two studies have indicated that in order to have sufficient power (.90) to detect cross-level interactions (that is, a level 2 slope relationship), it is necessary to have a sample of thirty groups containing thirty individuals each (Bassiri, 1988; van der Leeden & Busing, 1994).

There does appear to be a trade-off, however, between the number of groups and the number of individuals per group. For example, if there is a large number of groups (say, one hundred fifty), then the requirements regarding the number of individuals per group are reduced, to maintain the same level of power. Similarly, if there is a large number of individuals in each group, then the requirements regarding the number of groups will be reduced. All in all, however, there is still much work to be done regarding the power of these models and the sample size required in order for there to be adequate power (Bryk & Raudenbush, 1992).

**Conclusion**

In the preceding sections, we have provided a general introduction to and overview of hierarchical linear models. In an effort to provide a few summary conclusions regarding hierarchical linear models, we turn now to a discussion of the questions that we believe these models are both most and least effective at answering. In addition, we briefly compare hierarchical linear models to two commonly used methods of investigating multilevel models: OLS regression and within-and-between analysis. We close with some thoughts about what we believe the future holds for this approach to multilevel modeling.

**Questions That Hierarchical Linear Models Answer Most Effectively**

The following questions are particularly well suited to investigation by hierarchical linear models. The first four are related to one another and, in combination, build a picture of relationships across levels of analysis:

1. *Does the group in which individuals work make a difference?* This question concerns the degree to which individual measures vary across work groups. Hierarchical linear models address this question by estimating the proportion of total variance in individual responses that can be attributed to differences between groups and the proportion of total variance that can be attributed to differences among individuals. For example, the first step in the analysis of our substantive example was to investigate the extent to which the variance in helping behavior resided within groups versus between groups.

   Similar questions can be answered by a variety of ANOVA procedures that partition variance in individual responses (Shrout & Fleiss, 1979). However, hierarchical linear models also provide a framework for incorporating predictors of this variance at both the individual level and the work-group level. In this way, information provided about the degree of variation between work groups is a precursor to questions about the predictors of this variance. The influence of predictors at the individual level is the basis for the following question.

2. *What is the impact of individual differences across work groups?* This question concerns the nature of relationships in the level 1 units. The question about the impact of individual differences can be broken into two smaller questions. Returning to our substantive example, we see that the first question asks whether, on average, individual-level mood is related to helping behavior. As we
have seen, hierarchical linear models provide a pooled estimate of the overall impact of individual factors. The second question addresses the variability in this relationship across groups. This question is investigated by estimating the systematic variance in the slope of helping behavior when regressed onto mood. If there is variation across groups, we know that individual mood has different effects, depending on the work groups to which individuals belong.

3. Are individuals influenced by characteristics of the work group? This question asks whether particular characteristics of the work group are related to the variance in individual outcomes that resides between work groups. It is essentially a main-effect question, addressing whether group characteristics predict individual-level outcomes after controlling for individual differences (that is, level 1 predictors). Hierarchical linear models enable an assessment of the relative contribution of multiple work-group characteristics so that the unique proportion of variance in the individual outcome associated with a given work-group characteristic can be obtained.

4. Do properties of the work group modify individual-level relationships? Provided that the answer to the second question reveals significant variance in the level 1 relationship across groups, this question addresses the extent to which group-level predictors can account for this variance. The answer to this question is perhaps what constitutes the most distinct advantage that hierarchical linear models offer to organizational researchers. Organizations are often concerned with the impact of aggregate-level interventions on relationships at the individual level. For example, an incentive program may aim to influence the degree to which perceived fairness is related to contextual performance. A job design intervention may aim to increase the relationship between individual skills and task performance. In both examples, the relationship between two measures at the individual level is moderated by a characteristic of the work group. Hierarchical linear models are helpful here because its two-stage strategy enables assessment of the degree to which the relationship varies across work groups before an estimate is made of whether group characteristics moderate the relationship.

5. Do individual differences influence individual change over time? This question is somewhat different from the previous questions. It assumes that repeated measures have been obtained, and it asks whether the change in individuals over time is related to individ-

ual differences. Multiple measurements enable the assessment of specific growth patterns. For example, a linear and quadratic trend in work performance can be estimated for each individual. Subsequently, the researcher can investigate whether these trajectories are related to such individual differences as cognitive ability or previous work experience (Deadrick et al., 1997).

Questions That Hierarchical Linear Models Answer Least Effectively

The preceding five questions form the basis for a range of related questions that can be used to investigate multilevel relationships. Some questions that span levels of analysis are not readily answered by hierarchical linear models because they do not correspond to the assumptions made about the structure of the data and the types of relationships that are proposed. Three questions not addressed specifically by hierarchical linear models are described here:

1. How do individuals influence work-group characteristics? The difficulty of answering this question via hierarchical linear models is due to the assumption that the outcome measure is measured at the lowest level of analysis. Specifically, the question asks about the extent to which individual-level characteristics influence or predict aggregate phenomena. The examples discussed earlier were all constructed so that the dependent variable was measured at the lower level of analysis. When the outcome of interest is not at the lowest level, then this approach will not usually be the most appropriate analytical tool.

2. At what level of analysis should a variable be measured? Hierarchical linear models do not identify the most appropriate level at which a variable should be analyzed, although information obtained from the preceding question can shed light on the nature of the variation between work groups, which in turn can be used to support decisions about aggregation. For example, if a large proportion of variance in a measure occurs between work groups, then this result can be used to support aggregation of the measure to the work-group level. However, hierarchical linear models assume that the level of analysis for each measure has been determined before analysis, and that other techniques (for example, $r_{yg}$) are more appropriate for making decisions about level of measurement and data aggregation.
3. To which group should an individual be assigned? Hierarchical linear models require clear definition of the hierarchical structure in which measurements are embedded. The analysis does not provide information about the most appropriate group for an individual, as determined on the basis of ambiguous information about group membership. In this case, cluster-analytic procedures provide a more appropriate analytic tool.

Hierarchical Linear Models and OLS Regression

As already mentioned, OLS approaches present some problems with respect to multilevel modeling, but it is worthwhile to address several issues in greater depth. It will be recalled that OLS regression analog to this model is carried out by assigning the score on higher-level variables down to each of the respective lower-level units (for example, each member of a given group gets the same score on cohesion) and then conducting the analysis at the individual level. Given that OLS can approximate the hierarchical linear models, there are similarities in the sense that the same variables (both lower-level and higher-level) can be analyzed and the same types of relationships can be investigated (in-unit effects and both cross-level main effects and cross-level interaction effects).

This solution, however, is not without problems. OLS regression assumes, for example, that the random errors are independent, are normally distributed, and have constant variance. As Bryk and Raudenbush (1992) note, this assumption will likely be violated because the random errors will include a group-level component in addition to an individual-level component. This group-level error brings into question the independence of observations within groups because a portion of the random error is group random error (that is, measurement error of the group score), which is constant across individuals in a given group. In other words, the random errors of individuals in the same group are likely to be more similar than those of individuals in different groups, thus violating the assumption of independence. In addition, this group-level random error is also likely to vary across groups, thereby violating the assumption of constant variance (Bryk & Raudenbush, 1992). Finally, the assignment of group-level variables down to the individual level results in the use of statistical tests that are based on the number of individuals instead of on the number of groups. Therefore, standard errors associated with the tests of the group-level variable may be underestimated (Bryk & Raudenbush, 1992; Tate & Wongbundhit, 1983).

Hierarchical linear modeling, however, provides a more complex analysis and, in the process, more information than its OLS counterpart. Additionally, hierarchical linear modeling provides a more statistically appropriate analysis than OLS regression, for several reasons. First, hierarchical linear models explicitly partition the variance in the outcome variable and provide information about the magnitude (and significance) of these variance components. Second, separate regression analyses are performed for each group, relating the lower-level predictor(s) to the lower-level outcome. Because this is done, the level 1 intercepts and slopes are allowed to vary between level 2 units. OLS regression, by contrast, conducts a single regression analysis, pooling the lower-level units across groups, and subsequently does not allow the intercepts and slopes to vary. Fourth, as a result of the partitioning of the variance in the outcome into its within-group and between-group components, hierarchical linear modeling yields a more complex error term than its OLS counterpart. Specifically, the lower-level and higher-level errors are separately estimated, whereas the OLS regression approach combines them into a single term. Among other things, this has implications for assessing the explanatory power of variables at each of the different levels via the calculation of the R²'s (which is dependent on the error terms).

Hierarchical Linear Models and Within-and-Between Analysis

Another option for analyzing multilevel data is to use within-and-between analysis (Dansereau, Alutto, & Yammarino, 1984). In its basic form, WABA strives to answer two fundamental questions. The first question examines the variance within and between groups on a particular variable, in an effort to determine whether the variable is best represented at the individual or the group level. The second question examines the relationship between two variables, in an effort to determine whether the relationship resides primarily at the individual or the group level. It assumes that the two variables are measured at the same level of analysis—specifically, the
lower level. Although the basic WABA analysis is bivariate, it should be noted that Schriesheim (1995) has made the extension to a multivariate WABA.

In our view, the primary distinction between hierarchical linear models and WABA is the basic type of question that each technique was designed to answer. Hierarchical linear models are geared toward answering questions about how higher-level variables influence lower-level outcomes and lower-level relationships (that is, level 1 slopes). WABA is more geared toward answering questions about the level at which the relationship between two variables (both measured at the same level) resides. Thus, whereas WABA is concerned with identifying the level at which the relationship between two variables exists, hierarchical linear modeling prespecifies the level at which variables are expected to relate, and it specifies that their relationships will cross levels.

Another distinction, touched on earlier but made explicit here, is that WABA analyses involve only variables measured at the same level. For example, WABA can be used to investigate the relationship between two individual-level variables among individuals nested within groups. The analysis provides insight into whether the relationship between these two variables resides predominantly within groups, between groups, or both. Alternatively, hierarchical linear models allow for the investigation of the relationship between an outcome and variables measured and/or residing at two or more levels. For example, these models can be used to investigate, simultaneously, the influence of individual as well as team characteristics on individuals.

A final distinction between hierarchical linear models and WABA emerges when one is interested in investigating cross-level interactions. Hierarchical linear models were specifically designed to investigate how level 2 variables can moderate the relationship between two level 1 variables (that is, the slopes-as-outcomes model). WABA provides no such analysis and implicitly assumes that these slopes are homogenous across groups (George & James, 1993). Thus, although the WABA model has recently been expanded to include interactions at each level, cross-level interactions are not possible, to our knowledge. Although WABA does not enable a researcher to investigate one of the central questions for which hierarchical linear models were developed (that is, cross-level interactions), it does provide evidence surrounding questions that must be answered in advance of the application of hierarchical linear models (for example, the level at which a variable, or a relationship, resides).

The Future of Hierarchical Linear Models

At the beginning of this chapter, we noted the diverse fields in which hierarchical linear modeling has been applied. We believe that the popularity of these methods will continue to increase over the next five to ten years. In fact, we believe that the breadth of the fields in which these models have been adopted testifies to both the flexibility and the conceptual forthrightness of the method. Although levels-of-analysis issues have been discussed for a number of years in the organizational sciences (Rousseau, 1985), there seems to be a recent and growing emphasis on integrating levels of analysis into theory building (Cappelli & Sherer, 1991; House, Rousseau, & Thomas-Hunt, 1995; Klein et al., 1994; Morgeson & Hofmann, 1998). This volume speaks to this growing emphasis on integrating levels of analysis into organizational theory building. This increasing emphasis on multilevel theory building will result in a growing demand for multilevel data-analytic tools. We believe that hierarchical linear models will provide a mechanism for testing at least some of these theories (for example, those that specify predictors at multiple levels of analysis and an outcome variable at the lowest level).

Concurrent with this emphasis on multilevel theory, a number of applications of hierarchical linear modeling have started to appear in the organizational science literature. For example, in a special issue devoted to hierarchical linear models, the *Journal of Management* provided an overview of the technique (Hofmann, 1997), followed by four applications of these models to substantive questions (Deadrick et al., 1997; Griffin, 1997a; Kidwell et al., 1997; Vancouver, 1997). In addition, articles using hierarchical linear models have recently appeared in a number of other organizationally focused journals (Griffin, 1997b; Hofmann et al., 1993; Hofmann & Stetzer, 1996, 1998; Jimmieson & Griffin, 1998; Mosholder et al., 1998; Vancouver et al., 1994; Wech et al., 1998). The co-occurrence of this increasing demand for multilevel theory building and the emerging use of hierarchical linear models, as well as the use of other multilevel techniques described in this volume, will, we believe, create a synergy in the organizational sciences so that both
development and testing of multilevel models of organizational phenomena will greatly increase over the next five to ten years. We hope that this chapter, as well as the other chapters in this volume, will help to increase not only the strength of multilevel theories but also the degree to which these theories can be tested.

Notes
1. In some contexts, researchers are interested in obtaining the most accurate estimate for a level 1 parameter for a particular group. For example, often in educational research the level 1 parameters convey information about a particular school’s effectiveness (Raudenbush, 1988). The inspection of equations 1, 2, and 3, however, reveals that there are actually two estimates of these level 1 parameters. Specifically, one can obtain an OLS estimate of a particular group’s level 1 coefficients via equation 1, or use equations 2 and 3 to compute predicted values of a particular group’s intercept and slope. Thus, the OLS equation 1 provides one estimate of the intercept and slope, whereas the computation of predicted values in equations 2 and 3 provides a second estimate. If the researcher is interested in obtaining the most accurate estimate of the level 1 coefficients for a particular group, the question remains: Which of the two is best?

The HLM approach does not force the researcher to make a choice between these two different estimates of a unit’s intercept and slope. Specifically, the HLM procedure can produce empirical Bayes estimates, which are optimally weighted composites of the two different estimates discussed earlier—that is, the estimates emerging from equation 1 and those coming from equations 2 and 3 (Bryk & Raudenbush, 1992; Morris, 1983; Raudenbush, 1988). Both Bryk and Raudenbush (1992) and Raudenbush (1988) provide more extended discussions of these empirical Bayes estimates and their computation. Suffice it to say here that these empirical Bayes estimates can be computed and will provide more accurate estimates of a particular group’s level 1 coefficients, which in some contexts will be quite useful. Although it is important to point out that the ability to obtain these empirical Bayes estimates is an important strength of the HLM approach, they likely will not be relevant to most organizational researchers. Specifically, most organizational researchers are interested in obtaining level 2 predictors of the level 1 parameters instead of merely obtaining the best estimates of the level 1 parameters.

2. It should be noted here that 76 percent of the variance residing between groups is quite large when compared to published studies investigating cross-level models (e.g., Campion, Medsker, & Higgs, 1993; Hofmann & Stetzer, 1996; James, 1982). The reader is reminded that these are simulated data.

3. The sample-size requirements to detect a cross-level main effect (that is, the level 2 intercept model) are likely to be less, given that intercept estimates are typically more precisely estimated than are slope estimates (that is, they have small standard errors). In fact, Hofmann and Stetzer (1996) have found significant cross-level main effects by using data on twenty-one teams containing approximately ten individuals each.

References


James, L. R. (1995). Comments offered as part of a presentation (Introduction, explanation, and illustrations of hierarchical linear modeling as a management research tool) at the annual conference of the Academy of Management, Vancouver, British Columbia.


