2007

Hierarchical Linear Modeling in Organizational Research: Longitudinal Data Outside the Context of Growth Modeling

Irvin Sam Schonfeld
CUNY Graduate Center

David Rindskopf
CUNY Graduate Center

Hierarchical Linear Modeling in Organizational Research: Longitudinal Data Outside the Context of Growth Modeling

Irvin Sam Schonfeld,
City College of the City University of New York
The Graduate Center of the City University of New York
David Rindskopf
The Graduate Center of the City University of New York

Please address correspondence to Irvin Schonfeld, School of Education, City College of the City University of New York, NAC 6/207, New York, NY 10031 or ISchonfeld@ccny.cuny.edu. *We dedicate this paper to the memories of our fathers, George Schonfeld and Wallace Rindskopf.*

Organizational researchers, including those carrying out occupational stress research, often conduct longitudinal studies. Hierarchical linear modeling (HLM; also known as multilevel modeling and random regression) can efficiently organize analyses of longitudinal data by including within- and between-person levels of analysis. A great deal of longitudinal research has been conducted in the context of growth studies in which change in the dependent variable is examined in relation to the passage of time. HLM can treat longitudinal data, including data *outside the context of the growth study*, as nested data, reducing the problem of censoring. Within-person equation coefficients can represent the impact of Time $t - 1$ working conditions on Time t outcomes using all appropriate pairs of data points. Time itself need not be an independent variable of interest.

Keywords
hierarchical linear models; multilevel models; analysis of longitudinal data; methodology; occupational stress

The organizational research community has become aware of the utility of hierarchical linear modeling (HLM) in research (e.g., Bliese & Ployhart, 2002; Vancouver, Thompson, Tischner, & Putka, 2002; Vancouver, Thompson, & Williams, 2001). HLM (we use this term in the generic sense and not as an indication that we prefer one specific computer application over another) is also known as multilevel modeling, random regression, and random coefficients modeling. Several instructive publications on HLM are available, including those by Bliese and Jex (2002), Bliese and Ployhart (2002), Hox (2002), Raudenbush (1997), and Raudenbush and Bryk (2002). Our goal is to demonstrate a new use for these models in the context of longitudinal research. The new use of HLM, which strips away time as an independent variable of interest, contrasts with its application to growth modeling. Before turning to the new use, we briefly review key aspects of the application of HLM to growth studies.

Table 1: Interpretation of Notation in HLM Models for Longitudinal Research

Term	Interpretation
π_{0i}	Level-1, within-person intercept term for the *i*th person
π_{01}	Level-1, within-person intercept term for the 1st person
π_{02}	Level-1, within-person intercept term for the 2nd person
π_{03}	Level-1, within-person intercept term for the 3rd person
π_{1i}	Level-1, within-person coefficient of work environment for the *i*th person
π_{11}	Level-1, within-person coefficient of work environment for the 1st person
π_{12}	Level-1, within-person coefficient of work environment for the 2nd person
π_{2i}	Level-1, within-person coefficient of support for the *i*th person
π_{21}	Level-1, within-person coefficient of support for the 1st person
β_{00}	Level-2, between-person intercept term; the weighted average of all π_{0i} terms
β_{10}	Level-2, between-person coefficient of work environment; the weighted average of all level-1 π_{1i} terms
β_{20}	Level-2, between-person coefficient of support; the weighted average of all level-1 π_{2i} terms
β_{01}	Level-2, between-person coefficient of pre-employment *Y*, a factor that predicts the level-1, within-units intercept terms, the π_{0i} terms
β_{11}	Level-2, between-person coefficient of the multiplicative term representing the cross-level interaction of pre-employment *Y* and work environment
e_{ti}	The difference between the *Y* score predicted for the *i*th person at the *t*th time (given the level-1 and level-2 coefficients) and the score that person actually obtained
r_{0i}	The difference between the π_{0i} term of the *i*th person and β_{00}
r_{1i}	The difference between the π_{10} term of the *i*th person and β_{10}
r_{2i}	The difference between the π_{20} term of the *i*th person and β_{20}
var(e_{it})	The variance of the e_{it}, within-person terms across all individuals
var(r_{0i})	The variance of the r_{0i} terms, i.e., how the π_{0i} terms vary among people about β_{00}
var(r_{1i})	The variance of the r_{1i} terms, i.e., how the π_{1i} terms vary among people about β_{10}
var(r_{2i})	The variance of the r_{2i} terms, i.e., how the π_{2i} terms vary among people about β_{20}

Context of Growth Modeling

The concern of the growth model is change in individuals as a function of time. For illustrative purposes, we sketch a growth model in the context of a hypothetical 18-month study of growth in job knowledge among new entrants into work roles and the influence of a training component in adding to and accelerating job-related knowledge growth. Job knowledge is measured at job entry and at the conclusion of each of six quarters. All equations used in this article largely follow the notation employed by Raudenbush and Bryk (2001).

Before examining the effect of training, the investigator writes Equation 1, a level-1, within-person equation, and Equation 2, a level-2, between-person equation. Equations 1 and 2 are elements of the variance components model needed for ascribing variance in job knowledge to sources (a) within and (b) between individuals. We assume that the residual terms *e* and *r* are normally distributed with means of 0 and variances indicated by the general notation $\mathrm{var}(\cdot)$.

$$\mathrm{JobKnowl}_{ti} = \pi_{0i} + e_{ti} \quad (1)$$
$$\pi_{0i} = \beta_{00} + r_{0i} \quad (2)$$

Equations 1 and 2 set the stage for the calculation of the intraclass correlation coefficient (ICC), the ratio var(r)/[var(r) + var(e)]. In the present context, the ICC estimates the proportion of all variance in job knowledge that is between-person variance and is equivalent to "the average correlation between any pair of composite residuals" (Singer & Willett, 2003, p. 97), meaning the average autocorrelation. Table 1 summarizes the notation used in this and the next section.

Next, and still prior to an examination of the effects of training, the investigator examines the extent to which job knowledge grows, beginning with the individual's entry into the organization and continuing over the first 18 months of employment. The investigator rewrites the within-person equation such that it now includes time. Equation 3a includes time and time-squared terms and can be expanded to accommodate as many polynomial terms as the research question and the number of measurement occasions warrant. For ease of exposition, we created Equation 3b, a level-1 equation that does not extend beyond a linear term.

$$\mathrm{JobKnowl}_{ti} = \pi_{0i} + \pi_{1i}\mathrm{Quarters}_{ti} + \pi_{2i}\,\mathrm{Quarters}^2{}_{ti} + e_{ti}; \text{ where Quarters go from 0 to 6.} \quad (3a)$$
$$\mathrm{JobKnowl}_{ti} = \pi_{0i} + \pi_{1i}\mathrm{Quarters}_{ti} + e_{ti}; \text{ where Quarters go from 0 to 6.} \quad (3b)$$

The π_0 term in Equation 3b represents the amount of job knowledge the individual has at job entry, meaning zero quarters. The π_1 term represents growth in job knowledge per quarter on the job (or whatever other unit, e.g., weeks, months, years, etc., the researcher deems useful to gauge passage of time). An advantage of the HLM approach to growth data is that the spacing of the occasions at which job knowledge or other dependent variables are measured does not have to be uniform for all workers.

Equations 4a and 4b are level-2, between-person equations and suggest that the level-1 parameters, π_0 and π_1, vary across people. The variation of π_0 is about β_{00}, a weighted average of the π_0 terms, and variation of π_{1i} is about β_{10}, a weighted average of the π_1 terms. Equations 3b, 4a, and 4b are components of an unconditional growth model, representing time-related growth without regard to the presence of other factors that may influence either initial knowledge or rate of knowledge growth.

$$\pi_{0i} = \beta_{00} + r_{0i} \quad (4a)$$
$$\pi_{1i} = \beta_{10} + r_{1i} \quad (4b)$$

Estimates of the variance in the e_t, r_0, and r_1 terms are calculated and serve as a baseline against which variance estimates from models represented by more elaborate equations can be compared. One type of analysis that would be conducted at this stage is a test of whether the variances in the *r* terms in Equations 4a and 4b are nonzero. Comparing the deviances of models with and without a specific level-2 residual term, the investigator can ascertain whether variance in the r_0 and r_1 terms and the covariance between r_0 and r_1 differ significantly from zero. A test statistic having a chi-square distribution is obtained by subtracting from the deviance estimate (–2 × log likelihood) characterizing the less parameterized model[1] the deviance estimate characterizing the more parameterized model, a procedure amply described in several sources (e.g., Bliese & Ployhart, 2002; Hox, 2002; Raudenbush & Bryk, 2002; Singer & Willet, 2003).

All prospective workers are randomly assigned to traditional training or the training supplemented with a learning-to-learn (LTL) component. The LTL component is not only expected to increase job knowledge at job entry, but is also expected to carry forward into the workplace and increase the rate of on-the-job learning. Equation 5a, a between-person equation, represents the influence of the training component on initial knowledge. LTL is represented by a dummy variable set equal to 1 for job entrants who underwent the LTL component and 0 for job entrants who did not. Equation 5b represents the influence of LTL on the growth of knowledge over 18 months.

$$\pi_{0i} = \beta_{00} + \beta_{01}\,\mathrm{LTL}_i + r_{0i} \qquad (5a)$$
$$\pi_{1i} = \beta_{10} + \beta_{11}\,\mathrm{LTL}_i + r_{1i} \qquad (5b)$$

Equations 5a and 5b are substituted into Equation 3b to produce Equation 6, which explicitly indicates that both initial job knowledge, meaning knowledge at Time 0, and the slope of the line representing time-related change in job knowledge are affected by training.

$$\mathrm{JobKnowl}_{ti} = (\beta_{00} + \beta_{01}\,\mathrm{LTL}_i + r_{0i}) + (\beta_{10} + \beta_{11}\,\mathrm{LTL}_i + r_{1i})\,\mathrm{Quarters}_{ti} + e_{ti} \qquad (6)$$

The investigator can apply the *t* statistic to assess the hypothesis that the fixed effects β_{01} and β_{11} differ from zero (the hypothesis test that β_{00} is 0 is less interesting because β_{00} represents the predicted value of job knowledge when all the predictors are zero). Deviance statistics can be used to assess reduction in variance following the introduction of LTL. It is *not* our purpose to go into great detail on the application of HLM to growth models. We use growth models as a backdrop against which we examine the application of HLM to longitudinal data where growth is a minor issue or not an issue at all.

Longitudinal Data Analysis Outside the Growth-Study Context

Research on job knowledge lends itself to the application of growth modeling.

People acquire knowledge over time. Experiences such as training potentially contribute to knowledge growth over time. There are other job-related conditions that are not as time dependent. Occupational stress, an area of interest to many organizational researchers, is one of those conditions.

In longitudinal research, workers are followed over time, and data are collected over several observation periods. Ordinary least squares (OLS) regression and related procedures (e.g., repeated measures ANOVA) are problematic in the context of longitudinal research because of difficulty including in analyses individuals who continue, leave, and even reenter jobs. Standard errors produced by HLM when data are nested and applied in statistical tests of parameter estimates are more accurate than the standard errors produced by OLS. OLS and repeated measures procedures have difficulty integrating research participants who contribute data over differing lengths of time following the start of the study. By contrast, HLM is exceptionally well suited to research in which workers are followed longitudinally for varying lengths of time or at irregularly spaced intervals.

Vancouver, et al. (2001, 2002) applied HLM to analyses of data collected in four highly controlled, multitrial laboratory-based studies. For example, Vancouver et al. (2001) used performance on an analytical game in one trial to predict self-efficacy on the next trial. The (as many as) 10 trials, which the participants sequentially undertook over the course of (at most) an hour, represented a longitudinal dimension, albeit of only 60 minutes. Self-efficacy was *not* a function of the lateness of the trail. Trial number, a stand-in for time, was *not* employed as a predictor. Performance was the key predictor of self-efficacy.

We underline the utility of HLM in the context of longitudinal organizational research that can take place over months and years. HLM can provide a firm basis for analyzing longitudinal data bearing on the extent to which working conditions affect people who work. In such longitudinal research, "time need not be the independent variable" (Vancouver et al., 2001, p. 615).

An Application

Consider a study—whose hypothetical data on 355 social workers were specially generated for this article—in which an investigator follows a group of new social workers over time. One purpose of the study is to evaluate Dohrenwend and Dohrenwend's (1981) victimization model in the occupational-stress context. In this model, the accumulation of adverse work-related events (e.g., an episode of verbally abusive behavior initiated by a client; the physical assault of a coworker occurring at the workplace) increases the individual's risk of adverse psychological outcomes such as depressive symptoms. The study begins just before the social workers complete graduate school, 2.5 months before they obtain social work positions. During the pre-employment period, the investigator measures depressive symptoms, assuming that across-time carryover in pre-employment symptoms reflects something like trait distress (Schonfeld, 1996), which in this context will serve as a control variable. An alternative procedure would involve measuring negative affectivity (NA), or the propensity to experience dysphoric mood states (Watson & Clark, 1984). We return to this topic later in this article.

Because research on helping professionals suggests that work-related stressors exert relatively immediate effects on new job entrants (e.g., Schonfeld, 2001), the investigator, about 3 months after the social workers supply pre-employment data (and 2 weeks on the job), collects self-report data on working conditions and depressive symptoms. The investigator continues to collect such data every 3 months for the next 18 months. For ease of exposition, we label the pre-employment period Time 0 and the six data collection periods over the course of the next 11/2 years Times 1 to 6. Note that this study is *not* a growth study because the investigator is not concerned with the extent to which depressive symptoms rise or decline as a function of time (Bliese & Ployhart, 2002; Plewis, 1996).

The dependent variable is the Center for Epidemiologic Studies Depression scale (CES-D; Radloff, 1977). The investigator begins by developing a variance components model to estimate within- and between-person components of variance in depressive symptoms. Given that var(r) is 68.55 and var(e) is 62.86, the ICC is .52.

Although the focus of the study is the effect of accumulated adverse, work-related events on depressive symptoms, we do evaluate changes in the CES-D in relation to the passage of time. We find that the effect of time (β = .16, *SE* = .18) is nonsignificant. A test of time squared (β = –.10, *SE* = .14) is also nonsignificant. The effect of time is not central to the study. The core of the study lies in an examination of the influence of a time-varying covariate, or working conditions, on depressive symptoms.

A scale measuring working conditions is tailored for research on social workers and is administered every quarter the social worker is on the job. The investigator employs neutrally worded (Kasl, 1987) scale items that assess the frequency of adverse working conditions (e.g., How often have you observed a fight at the center? 0 = *not at all*; 1 = *once per month*; 2 = *once per week*; 3 = *two to four times a week*; 4 = *daily*), insult from clients, and so on. The work environment variable is a time-varying predictor because working conditions change over time (e.g., the number of multiproblem clients fluctuates with time). The investigator also examines another important time-varying factor, coworker support, a variable thought to reduce psychological distress. In developing an analytic plan for the study, the investigator writes Equation 7, a level-1, within-person equation. The investigator also writes three level-2, between-person equations, Equations 8a, 8b, and 8c. Each β term is a level-2 coefficient representing a weighted average of the corresponding π terms. Variation in each level-2 r term represents variability in the corresponding level-1 π term.

$$\text{CES-D}_{ti} = \pi_{0i} + \pi_{1i}\text{WorkEnv}_{(t-1)i} + \pi_{2i}\text{Support}_{(t-1)i} + e_{ti} \quad (7)$$

$$\pi_{0i} = \beta_{00} + r_{0i} \quad (8a)$$

$$\pi_{1i} = \beta_{10} + r_{1i} \quad (8b)$$

$$\pi_{2i} = \beta_{20} + r_{2i} \quad (8c)$$

In this model, for each social worker, there are up to five sets of data points that contribute to estimating parameters in Equation 7. The purpose of the equation is to predict the CES-D at Time t from working conditions (and coworker support) at Time $t - 1$. Among social workers who participated at every data-collection period, the Time 2 CES-D is paired with the Time 1 work environment scale (and support); the Time 3 CES-D with Time-2 work environment (and support); the Time 4 CES-D with the Time 3 work environment; the Time 5 CES-D with the Time 4 work environment; and the Time 6 CES-D with the Time 5 work environment. Thus the within-person, level-1 prediction equation is generated from these five sets of data points; one equation is estimated from the five data points. One equation is created for each person, although HLM is flexible enough to generate a within-person equation for workers who are not present for every data-collection period (albeit with larger standard errors for the level-1 coefficients generated). Note that time is part of the subscript in the within-person equation (the subscript marking the occasion depressive symptoms are measured); time, however, is not a level-1 predictor here as it is in the growth models seen earlier. Equation 7 suggests that the intensity of a person's level of depressive symptoms can vary over time; it does not have to increase or decrease in a steady manner as a function of time.

The predictors in Equation 7 are centered. Work environment is centered about the value 1, which is meaningful in the sense that the score is a benchmark reflecting bad things happening (e.g., observing a fight at a social work center) at a rate of once per month. One is also the modal value for the items comprising the scale and very close to the scale mean ($M = 1.13$). Coworker support is a measure based on items (e.g., How much can your fellow social workers be relied on when things get tough at work? 1 = *not at all*,2 = *a little*, 3 = *somewhat*, 4 = *very much*) adapted from House's (1980) coworker support scale. The scale is centered about 3, the modal response to the individual items. The score is also close to the scale's mean ($M = 3.18$). One advantage of centering both scales at values close to their respective means accrues when testing for an interaction. Centering reduces collinearity between an interaction term and the scales that are multiplied to create the interaction term.

Equations 8a, 8b, and 8c represent between-person equations. At this level (and consistent with our discussion of growth models), each r term is treated as a normally distributed random variable, with the β term being the weighted mean of the corresponding π_{ij}s. The key findings are in Table 2.

The term β_{00} represents an average CES-D score, across all individuals and occasions, and adjusted such that all predictors are set to zero. The score of an “average” novice social worker is β_{00}, and β_{00} differs significantly from zero. The adjusted mean of the CES-D score, β_{00}, when compared to mean scores from general population samples (Schonfeld, 1990), is somewhat elevated, suggesting that being a novice social worker is stressful.

Table 2: The Hierarchical Linear Modeling Coefficients and Variances for the Equations in Which Y Was Regressed on the Level-1, Within-Person Factors, Work Environment and Support

	Coefficient	Effect size	SE	p	var(*r*)	p
Intercept	β_{00}	14.31	0.47	.001	48.49	.001
Work environment	β_{10}	3.66	0.72	.001	10.92	.05
Support	β_{20}	–1.74	0.60	.01	[a]	
					var(e_{it})	
					63.92	

Note: Using the restricted iterative generalized least square approach to model fitting, we examined the level-2 variances by computing the differences in the deviance statistics between various models with and without a particular *r* term.

a. The variance in the r_{2i} terms and the covariances between r_{2i} and the r_{0i} and r_{1i} terms do not differ significantly from zero. The above model shows the fixed and random effects when the r_{2i} term was deleted from the model.

As indicated by *t* statistics, the β_{10} and β_{20} terms differ significantly from zero. A unit increase on the work environment scale is associated with a 3.7-point increase in the CES-D. A unit increase in coworker support is associated with a 1.7-point decrease in symptoms. Using a restricted iterative generalized least squares approach to the stochastic components of our models, a comparison of deviance statistics (involving a baseline model that contains the r_0 term and the two predictors) indicates that we cannot delete the r_{1i} variance and the covariance between that term and the r_{0i} term ($\chi^2[2] = 6.29$; $p < .05$); however, the r_{2i} variance and the covariance involving that term and the r_{0i} term do not differ significantly from zero ($\chi^2[2] = 4.14$). The investigator retains the r_{1i} term but not the r_{2i} term. These findings are compatible with the view that there are individual differences in the impact of the work environment on symptoms but not in the impact of coworker support. Caution is required in deciding to delete a residual term. It is possible that small sample size (not the case here) and consequent lack of power may be the basis for nonsignificant variance in a residual term, in which case it would be unwise to delete the term.

As indicated earlier, the propensity to experience psychological distress can affect an outcome such as depressive symptoms. It can also influence predictors such as the work environment scale by affecting how the incidents that make up the scale (e.g., an encounter with an angry client) are remembered. In addition to creating reporting biases, high levels of distress can also affect an individual’s comportment and thereby affect coworkers' willingness to offer support. We employ pre-employment depressive symptoms "as an indicator of an established disposition toward psychological distress that is likely to be strongly related to neuroticism and negative affectivity" (Lennon, Dohrenwend, Zautra, & Marbach, 1990, p. 1044). The pre-employment depressive symptoms scale, although susceptible to the influence of pre-employment nonwork stressors, for the purpose of this study, is conceptualized as a time-invariant factor[2] representing initial severity (cf. Gibbons et al., 1993) as well as the individual's propensity to experience psychological distress. Given the nonzero variance in the r_{1i} terms, pre-employment symptoms may also account for some of the between-person variability in the impact of the work environment (variability in the π_{1i}s). Pre-employment symptom level is a factor that requires control. Similarly, the individual's pre-employment blood pressure represents a factor that should be controlled when studying the impact of working conditions on blood pressure following entrance into a work role. Pre-employment blood pressure represents an initial level of severity that influences later blood pressure measurements and is the platform from which working conditions begin to exert their effects. In view of the importance of controlling pre-employment levels of outcomes, Equations 8a and 8b have been modified to create Equations 9a and 9b, respectively, by including pre-employment CES-D. We do not add pre-employment CES-D to Equation 8c because analyses described above indicate that the r_{2i} variance does not differ significantly from zero. There is no need to account for variance in r_{2i}.

$$\pi_{0i} = \beta_{00} + \beta_{01}\text{Pre.CES-D}_i + r_{0i} \quad (9a)$$
$$\pi_{0i} = \beta_{10} + \beta_{11}\text{Pre.CES-D}_i + r_{1i} \quad (9b)$$

The variance estimates of r_{0i} and r_{1i} are now conditional estimates, conditioned on the influence of the level-2, between-person predictor pre-employment CES-D. Equations 9a and 9b are substituted into Equation 7 (with the r_2 term deleted), creating Equation 10, revealing how pre-employment levels of Y influence during-employment levels.[3]

$$\text{CES-D}_{ti} = (\beta_{00} + \beta_{01}\ \text{Pre.CES-D}_i + r_{0i}) + (\beta_{10} + \beta_{11}\ \text{Pre.CES-D}_i + r_{1i})\text{WorkEnv}_{(t-1)I} + \beta_{20}\text{Support}_{(t-1)i} + e_{ti} \quad (10)$$

We found that var(r_{0i}), although considerably smaller than in the previous model, remains significantly different from zero (see Table 3)—var(r_{1i}) is now larger, a phenomenon that is unlikely to happen in OLS regression but is part of the landscape in HLM (see Singer & Willet, 2003). The investigator potentially can account for individual differences among the π_{0i}s by including other between-person covariates (e.g., age at entrance into the social work profession) that may further reduce the variance in the r_{0i} term. The introduction of pre-employment CES-D (whether with or without the interaction with the work environment) reduces the effect of Coworker Support. The interaction of pre-employment CES-D and working conditions was marginally significant.

The analyses could continue with the addition of a within-person interaction (Work Environment × Coworker Support). We stop at this juncture because the illustrative example is sufficient to provide an idea of how to proceed in applying HLM without time and growth being the central focus.

Table 3: The Hierarchical Linear Modeling Coefficients and Variances for the Equations in Which Y Was Regressed on the Within-Person Factors, Work Environment and Support, and the Between-Person Factor Pre-Employment Y

	Coefficient	Effect size	*SE*	*p*	var(*r*)	*p*
Intercept	β_{00}	14.24	0.43	.001	26.70	.001
Work environment	β_{10}	3.86	0.64	.001	16.60	.01
Support	β_{20}	–1.11	0.57	.06	[a]	
Pre-employment Y	β_{01}	0.45	0.04	.001		
Pre-employment Y × work environment	β_{11}	0.12	0.07	.09		
					var(e_{it})	
					41.19	

Note: The variables in *italics* are level-2, between-person variables; the non-italicized variables are level-1 variables.
a. As per Table 2, the variance and covariances associated with the r_{2i} term do not differ significantly from zero, and the term has been deleted from the model.

Autocorrelation and Heteroscedasticity

The autocorrelation in the above model is a conditional autocorrelation (Singer & Willett, 2003), conditioned on the influence of working conditions and coworker support. It could be smaller, or greater, than the autocorrelation in the unconditional variance components model. Most software for multilevel models allows various structures for such autocorrelations to be specified and tested. Although it could be difficult to eliminate autocorrelations, our approach is to introduce time-varying predictors that will reduce those autocorrelations. Autocorrelations may result from incomplete model specification. However, it is not always possible to measure key omitted variables, in which case the usual method of modeling the remaining autocorrelation will be necessary.

It should be stressed that the parameter estimates that are of most interest to investigators are typically the estimates of βs, meaning the fixed effects. Raudenbush and Bryk (2002) found that parameter estimates of fixed effects are not biased even if the investigator gets the exact structure of correlated error wrong. Furthermore, for a variety of error covariance structures (with the possible exception of the compound symmetry model) the standard errors for the βs may not be seriously biased. Heteroscedasticity can derive from "the effects of omitted predictors" (pg. 84 Singer & Willett, 2003) although it is also possible that it reflects real effects, meaning some processes may result in heterogeneous variances over time. "The homogeneity assumption is not per se a serious problem for estimating either level-2 coefficients or their standard errors" (Raudenbush & Bryk, 2002, p. 264). In the HLM context, heteroscedasticity is not the problem it is in the context of OLS regression. Variability in variances can be built into HLM models (e.g., Woodhouse, Rasbash, Goldstein, Yang, & Plewis, 1996). Current software allows the investigator to evaluate variable-related changes in residual variance.

Power

A design issue in any study is choosing a sample size to obtain sufficient power to detect effects. In the context of the application of HLM to longitudinal data sets, power is a function of the number of persons, the number of observation occasions per person, and the amount of within- and between-person variation. Cohen (1998) and Raudenbush (1997; Raudenbush & Liu, 2000, 2001) provide more complete coverage of the topic. We also direct the reader to *Optimal Design*, power analysis software developed by Raudenbush and his colleagues (Raudenbush & Congdon, 2001; Raudenbush, Liu, Congdon, & Spybrook, 2004) for HLM applications.

Missing Data

Missing data are not as problematic when using HLM methods as they are with traditional methods that assume the same number of observations, and at the same time points, for all people. However, there are still potential problems with most methods of analysis if data are missing. Modern missing data methods classify missing data into one of three types: missing completely at random (MCAR), (2) missing at random (MAR), and (3) nonignorable nonresponse (Little & Rubin, 1987). Data are MCAR if the missing values are unrelated to any variables in the data set. Data are MAR if missingness in a variable is not a function of the missing value (although the missing value can be a function of another variable). Nonignorable nonresponse occurs if the missing data are related to the value that would have been observed, after controlling for variables that are not missing. For example, if a potentially high CES-D score were missing because a worker was hospitalized for depression, it is almost certain that we have a case of nonignorable nonresponse. Models for nonignorable nonresponse are generally complicated and depend on untestable assumptions; such models are not often fitted in practice. The HLM methods we are discussing can easily handle data that are MAR; HLM is not limited to the more restrictive case of MCAR.

Concluding Remarks

HLM is a highly flexible and powerful tool that is especially suited for analyzing the kinds of longitudinal data organizational researchers collect. We want to highlight the special strength of HLM in longitudinal research. It has *more* to offer than as a tool for analyzing growth data (see Rindskopf & Wallen, 2003; Vancouver et al., 2001, 2002). HLM handles multiwave studies in which time is not a variable of interest. It is especially suited for research in contexts in which work characteristics and outcomes are measured over several data-collection periods. HLM handles participant loss with a minimum of censoring. However, we emphasize that the timing of data collection periods must be predicated on a solid theory of the processes involved as well as on findings from past research. Although HLM can integrate longitudinal data from participants who participate at nonuniform time points (e.g., one participant participates at 3.3 months, 9.5 months, and 15.2 months from Time 0; another, at 5.8 months, 12.3 months, and 21.7 months from Time 0), HLM cannot patch up a longitudinal study in which the timing of data collection is out of step with the reality of the processes under study.

A variety of software programs is available. These include MLwiN1.10 (Goldstein et al., 1998), the application software known as HLM 6 (Raudenbush, Bryk, Cheong, & Congdon, 2004), and Mixreg (Hedeker & Gibbons, 1996a), to cite just three. The investigator may consider applying HLM procedures to research involving dichotomous outcomes such as disease endpoints. Caution, however, is warranted when considering dichotomous outcomes because some earlier methods have produced biased results (see Rodriguez & Goldman, 2001). MLwiN1.10, HLM 6, and companion software to Mixreg called Mixor (Hedeker & Gibbons, 1996b) can be applied to research involving dichotomous nominal outcomes. Moreover, Mixor is additionally adapted for ordinal regression models. Both MLwiN1.10 and HLM 6 can be applied to data sets having multicategory outcomes. Both can also be applied to multilevel count outcome data (multilevel Poisson data).

The matter of whether to use a growth model is not an either–or proposition. There are occasions when investigators will not have sufficient information to model differences within individuals among observational periods, in which case time will suffice. We urge investigators to look widely at the landscape of models appropriate to their research questions and reach beyond models in which time alone is the within-person independent variable. Sometimes investigators need both time and a variety of other independent variables. There can of course be growth (e.g., in the case of learning and fatigue). But we suspect the biggest payoff will come from looking toward person-level independent variables.

Notes

1. The deviance statistic is actually equal to –2 × (log likelihood of the model – log likelihood of the saturated model). However, the log likelihood of the saturated model is equal to 0 (see Singer & Willett, 2003, p. 117).
2. A pre-employment measure of depressive symptoms does not technically represent the type of time-invariant, between-person factor that investigators treat as a level-2 variable whereas NA, at least in theory, constitutes relatively stable personality trait. In practice, however, NA measures may be affected by stressors and mood and change over time (Spector, Zapf, Chen, & Frese, 2000).
3. An alternative approach to the analysis would involve the following level-1 equation: $\text{CES-D}_{ti} = \pi_{0i} + \pi_{1i}\,\text{WorkEnv}_{(t-1)i} + \pi_{2i}\,\text{Support}_{(t-1)i} + \pi_{3i}\text{CES-D}_{(t-1)i} + e_{ti}$. In this approach, depressive symptoms at Time *t* are regressed on working conditions, support, and depressive symptoms at Time $t - 1$. The

argument favoring this approach acknowledges that measures of depressive symptoms are likely to be serially correlated and that working conditions at Time $t - 1$ are likely to be correlated with contemporaneous symptoms, producing a spurious relation between Time t symptoms and Time $t - 1$ working conditions. With Y at Time $t - 1$ now controlled *along with* pre-employment Y, the effects associated with the work environment are reduced.

The approach to data analysis described in the text of the article holds that controlling for the level-1 factor depressive symptoms at Time $t - 1$ represents a kind of "overcontrol" when pre-employment symptoms are controlled as a level-2 factor. This is because the Time $t - 1$ CES-D is likely to be affected by Time $t - 1$ working conditions (Schonfeld, 1996, 2001). The purpose of controlling Time 0 symptoms is to control the carryover of psychological distress across time. The carryover represents trait effects. Two factors will largely affect depressive symptoms at Times 1, 2, 3, and so on, making them ill suited as level-1 control variables. One is the background trait carryover. Symptoms at any time period *after* Time 0 are affected by trait depression/distress or NA. Trait distress/NA is already controlled in the level-2 equation. The other factor is the work stressors, not to mention support. Time $t - 1$ symptoms are about as likely as Time t symptoms to be affected by Time $t - 1$ working conditions, especially if the lag is relatively brief as it is in the present study. Controlling for Time $t - 1$ symptoms will thus likely obscure the influence of working conditions on later symptoms. If the investigator is already controlling for trait effects in the form of level-2 Time-0 depressive symptoms or NA, there may not be the need to introduce $Y_{(t-1)}$ as a control variable in the level-1, within-person equation.

As more multiwave longitudinal studies accumulate, particularly studies that begin during a pre-employment period to gain the advantage of taking baseline measurements of the dependent variable, more on the subject of identifying suitable control variables is likely to be heard.

References

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.

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustration. *Organizational Research Methods, 5,* 362-387.

Cohen, M. (1998). Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics, 14,* 267-275.

Dohrenwend, B. S., & Dohrenwend, B. P. (1981). Life stress and illness: Formulation of the issues. In B. S. Dohrenwend & B. P. Dohrenwend (eds.), *Stressful life events and their contexts* (pp. 1-27). New York: Prodist.

Gibbons, R. D., Hedeker, D., Elkin, I., Waternaux, C., Kraemer, H. C., Greenhouse,

J. B., et al. (1993). Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry, 50,* 739-750.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., et al. (1998). *A user's guide to MLwiN.* London: Multilevel Models Project, Institute of Education, University of London.

Hedeker, D., & Gibbons, R. D. (1996a). Mixreg: A computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine, 49,* 229-232.

Hedeker, D., & Gibbons, R. D. (1996b). Mixor: A computer program for mixed-effects ordinal regression. *Computer Methods and Programs in Biomedicine, 49*, 157-176.

House, J. S. (1980). *Occupational stress and the mental and physical health of factory workers*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Kasl, S. V. (1987). Methodologies in stress and health: Past difficulties, present dilemmas, future directions. In S. V. Kasl & C. L. Cooper (Eds.), *Stress and health: Issues in research methodology* (pp. 307-318). Chichester, UK: Wiley.

Lennon, M. C., Dohrenwend, B. P., Zautra, A. J., & Marbach, J. J. (1990). Coping and adaptation to facial pain in contrast to other stressful life events. *Journal of Personality and Social Psychology, 59,* 1040-1050.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Plewis, I. (1996). Statistical methods for understanding cognitive growth: A review, a synthesis, and an application. *British Journal of Mathematical and Statistical Psychology, 49,* 25-42.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2,* 173-185.

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y., & Congdon, R. T. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.

Raudenbush, S. W., & Congdon, R. (2001). *Optimal design. Version 0.19.* Lincolnwood, IL: Scientific Software International.

Raudenbush, S. W., & Liu, X. (2000). Statistical analysis and optimal design for multisite randomized trials. *Psychological Methods*, *5,* 199-213.

Raudenbush, S. W., & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6,* 387-401.

Raudenbush, S. W., Liu, X., Congdon, R., & Spybrook, J. (2004). *Optimal design for longitudinal multilevel research: Documentation for the "Optimal Design"* *software*. Ann Arbor, MI: University of Michigan.

Rindskopf, D. & Wallen, A. S. (2003). *Analysis of time-independent repeated measures data using multilevel models*. Manuscript submitted for publication.

Rodriguez, G., & Goldman, N. (2001). Improved estimation procedures for multilevel models with binary response: A case study. *Journal of the Royal Statistical Society, Series A, General, 164,* 339-355.

Schonfeld, I. S. (1990). Psychological distress in a sample of teachers. *Journal of Psychology, 124,* 321-338. Schonfeld, I. S. (1996). Relation of negative affectivity to self-reports of job stressors and psychological outcomes. *Journal of Occupational Health Psychology, 1,* 397-412.

Schonfeld, I. S. (2001). Stress in 1st-year women teachers: The context of social support and coping. *Genetic, Social, and General Psychology Monographs, 127,* 133-168.

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

Spector, P. E., Zapf, D., Chen, P. Y., & Frese, M. (2000). Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior, 21,* 79-95.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87,* 506-516.

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86,* 605-620.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465-490.

Woodhouse, G., Rasbash, J., Goldstein, H., Yang, M., & Plewis, I. (1996). *Multilevel modeling applications: A guide for users of MLn.* London: Multilevel Models Project, Institute of Education.

Irvin Sam Schonfeld, School of Education, City College of the City University of New York and departments of psychology and educational psychology, the Graduate Center of the City University of New York.

David Rindskopf, department of educational psychology, the Graduate Center of the City University of New York.