Researchers commonly ask whether relationships between exogenous predictors, X, *and outcomes,* Y, *are mediated by a third set of variables,* Z. *Simultaneous equations decompose the relationship between* X *and* Y *into an indirect component, operating through* Z, *and a direct component, the relationship between* X *and* Y *given* Z. *Often,* X, Y, *and/or* Z *are measured with error. Structural equation modeling is widely used in this scenario. However, sociological data commonly have a nested structure (students within schools, residents within local areas). Hierarchical linear models represent such multilevel data well and can handle errors of measurement, but have not incorporated simultaneous equations for direct and indirect effects. This article incorporates the study of such mediated effects into the hierarchical linear model, naturally extending the analysis to include unbalanced, multilevel designs and missing data. The authors illustrate the approach by examining the extent to which neighborhood social control mediates the relationship between neighborhood social composition and violence in Chicago.*

Assessing Direct and Indirect Effects in Multilevel Designs With Latent Variables

STEPHEN W. RAUDENBUSH
University of Michigan
ROBERT SAMPSON
University of Chicago

Of central interest in social science are hypotheses about mediators of well-known relationships. Researchers ask whether cognitive skills mediate the effect of education on occupational status (Rivera-Batiz 1992; Bowles and Gintis 1996), whether disciplinary climate mediates the link between school social composition and student achievement (Lee and Bryk 1989), and whether birth control practices mediate the effect of maternal education on fertility

AUTHORS' NOTE: *This research was supported by the Project on Human Development in Chicago Neighborhoods (PHDCN) with funding from the MacArthur Foundation and the National Institute of Justice. Thanks are due to Felton Earls, Albert Reiss, and Steven Buka for their essential role in conceptualizing the PHDCN and to Richard Congdon for the programming of all computations. Thanks are also due to anonymous reviewers for important suggestions. Matheos Yosef checked the derivations and computed all results.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 28 No. 2, November 1999 123-153

(Mason, Wong, and Entwisle 1984). Structural equation modeling (SEM) is widely used in such studies (Yuan and Bentler 1997; Jöreskog and Sörbom 1996; Muthen 1990). These models readily incorporate simultaneous equations that represent mediated effects (Bollen 1989),[1] and they provide a convenient framework for incorporating information about errors of measurement of predictors, mediators, and outcomes. Maximum likelihood (ML) provides simultaneous and statistically efficient estimation of all direct and indirect effects among the latent variables, that is, the variables measured with error.

The study of mediated effects often requires multilevel data. In Lee and Bryk's (1989) study, social composition and disciplinary climates varied at the school level whereas the outcome varied at the student level. In violence prevention experiments, social units such as schools or neighborhoods are assigned randomly to treatments, and researchers ask whether observed treatment effects on student behavior are mediated through measurable aspects of treatment implementation (Powell et al. 1996). Hierarchical linear models are widely applicable in such multilevel settings, and, as is shown in research reviewed below, methodologists have recently made significant progress incorporating measurement errors into these models. However, to date, the literature on hierarchical linear models has not addressed the problem of estimating mediated effects.

The innovation of this article is to incorporate mediational models for latent variables into the hierarchical linear model. The advantage of this approach is that the analyst can easily incorporate a variety of unbalanced, multilevel data structures, for example, data on students within schools, workers within firms, or residents within census tracts. In the illustrative example, we show how to tailor the model to incorporate commonly arising difficulties such as item-level missing data and response bias. Heteroscedastic data and random coefficients are readily handled by the approach.

LATENT VARIABLES IN MULTILEVEL DESIGNS

Adapting latent variable analysis for multilevel designs currently constitutes a vigorous line of methodological research. One stream of inquiry incorporates adjustments for measurement error in hierarchical regression models. Thus, Longford (1993) provided ML estimation for two-level regression models involving latent variables, each having multiple indicators. Goldstein (1995) derived adjustments to generalized least squares regression that correct for measurement errors in the predictors, assuming prior knowledge of the measurement error variances. And Woodhouse et al. (1996) assessed the consequences for statistical inference of appropriate adjustments for measurement errors at each of two levels.

A second stream of inquiry aims to generalize mean and covariance structure modeling via SEM to incorporate multilevel data. McDonald and Goldstein (1989) and Lee (1990) provided the needed theory for ML estimation for unbalanced two-level models. Muthen (1990) showed that currently available software for SEM can provide ML estimates in the case of two-level data having balanced designs (equal numbers of level-1 units within each level-2 unit). McDonald (1993, 1994) developed a more general approach that allows missing data at either level, implemented by specialized software. Raudenbush (1995) showed how to construct an expectation-maximization (EM) algorithm for ML estimates in unbalanced designs, relying on iterative reestimation using standard SEM software. An important goal of this stream of inquiry is to extend the wide range of covariance structures of SEM to the multilevel setting. It appears difficult, however, to extend the approach beyond two levels or to include random coefficients without substantially novel algorithms and software.

We adopt a third approach: to represent measurement error as a level within the hierarchical model. Raudenbush, Rowan, and Kang (1991) represented the item responses in a survey as level-1 units in a three-level model. The level-1 design matrix linked the items with the "true scores" or latent variables being measured for each teacher in the study. At level 2, the latent variables varied among teachers sampled within schools. And at level 3, the coefficients of the level-2 model varied randomly across schools. In essence, then, the level-1 model was a measurement model, whereas the next two levels represented a multivariate hierarchical linear model for the latent variables. The approach easily handles missing data, multiple levels of nesting, and random coefficients, and can be implemented with software that is now widely available.[2]

TASKS OF THIS ARTICLE

In this article, we extend the approach of Raudenbush et al. (1991) to study mediated effects involving latent variables. We present the underlying theory, illustrate how to specify the model, and examine the consequences of ignoring measurement error in the context of an example. We show how item nonresponse and survey response biases can be incorporated in the model. We also link the study of direct and indirect effects to the comparison of coefficients between models (Clogg, Petkova, and Haritou 1995; Allison 1995).

To illustrate our approach and its extensions, we consider a seminal and longstanding relationship in the criminological literature—that between neighborhood social composition and rates of criminal violence. The basic hypothesis is that the well-established association between social composition (poverty concentration, ethnic isolation, and percent foreign born) and levels of violence in urban neighborhoods is largely mediated by neighborhood informal social control (Sampson and Lauritsen 1994). In particular, it has been argued that concentrated poverty undermines the capacity for informal social control, in turn increasing the risk of violence. In this example, informal social control is a neighborhood-level predictor measured by combining the item responses of residents within each neighborhood. The data present difficulties that we believe arise commonly in multilevel survey research, including item missing data, possible response bias, and unbalanced design. As the example illustrates, we find the hierarchical model to be a natural framework for coping with these difficulties.

Subsequent sections present the model, derive estimators, and consider the illustrative data, showing in detail how to use current software to implement the approach. We close by considering the current strengths and limitations of the approach and formulating questions for further inquiry.

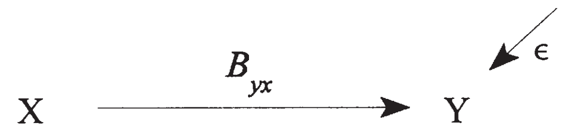

Figure 1: Total Effect of *X* on *Y*

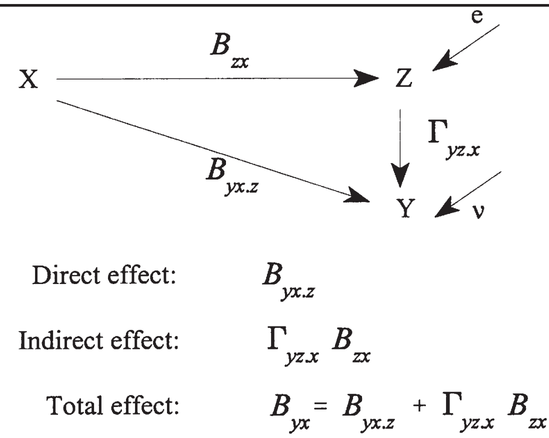

Figure 2: Direct Effect of *X* on *Y*, Given *Z*, and Indirect Effect of *X* on *Y*, Through *Z*

THE MODEL

STRUCTURAL MODEL

Our attention is confined to a set of exogenous predictors, *X*, outcomes, *Y*, and possible mediators, *Z*. In most applications, linear associations between *X* and *Y* will be well established, and we are interested in studying whether *Z* mediates or explains this relationship. This scenario is depicted in Figures 1 and 2. Figure 1 represents the *total* effect of *X* on *Y*, quantified by the (unstandardized) regression coefficient matrix B_{yx}. Figure 2 incorporates *Z* as a potential mediator. If the paths between *X* and *Z* (as indicated by B_{zx}) and between *Z* and *Y* (characterized by $\Gamma_{yz.x}$) are nonzero, we say that there is an *indirect* effect of *X* on *Y* through *Z*. If $B_{yx.z}$ is nonzero, we say there is a *direct* effect of X on Y, controlling Z. And if the indirect effect is nonzero, we commonly conclude that Z partially *mediates* the effect of X on Y.

Figure 2 can be represented by a pair of simultaneous equations:

$$Y_k = \mathbf{B}_{yx.z} X_k + \Gamma_{yz.x} Z_k + \blacklozenge_k \qquad \blacklozenge_k \sim N(0, \Sigma) \tag{1}$$

$$Z_k = \mathbf{B}_{zx} X_k + e_k \qquad e_k \sim N(0, \mathbf{T}_{zz}). \tag{2}$$

In equation (1), Y_k is the $p \times 1$ vector of outcomes; X_k is the $r \times 1$ vector of exogenous predictors, and $\mathbf{B}_{yx.z}$ is the $p \times r$ matrix of coefficients associated with them; Z_k is the $q \times 1$ vector of mediating variables, and $\Gamma_{yz.x}$ is the $p \times q$ matrix of associated regression coefficients. Equation (2) similarly represents the linear association between X and Z. Substituting equation (2) into equation (1) gives the reduced form of the model,

$$\begin{aligned} Y_k &= (\mathbf{B}_{yx.z} + \Gamma_{yz.x} \mathbf{B}_{zx}) X_k + \Gamma_{yz.x} e_k + \blacklozenge_k \\ &= \mathbf{B}_{yx} X_k + \varepsilon_k . \end{aligned} \tag{3}$$

Equation (3) reveals the usual decomposition of the total effect, $\mathbf{B}_{yx} = \mathbf{B}_{yx.z} + \Gamma_{yz.x}\mathbf{B}_{zx}$ into the sum of the direct effect, $\mathbf{B}_{yx.z}$, and the indirect effect, $\Delta = \Gamma_{yz.x}\mathbf{B}_{zx}$. We note that the indirect effect is equivalent to the difference between the coefficients for X in the reduced model (equation (3)) and in the full model (equation (1)), a fact that is useful in constructing tests of comparisons between models as shown in Clogg et al. (1995) and Allison (1995). Assuming the model given by equation (3) to be accurate also imposes structure on the variance-covariance matrix of the reduced model, specifically:

$$Var(Y_k \mid X_k) = \mathbf{T}_{yy} = \Gamma_{yx.z} \mathbf{T}_{zz} \Gamma^{T}_{yx.z} + \Sigma. \tag{4}$$

MEASUREMENT MODEL

Often, some or all elements X, Z, and Y are measured with error, and one must attempt to estimate the parameters of equations (1) and (2) from fallible data. We therefore formulate a simple measurement model,

$$\begin{aligned} Y_{obs_k} &= D_{yk} Y_k + e_{yk} \\ X_{obs_k} &= D_{xk} X_k + e_{xk} \\ Z_{obs_k} &= D_{zk} Z_k + e_{zk}. \end{aligned} \tag{5}$$

Here, $Y_{obs_k}, X_{obs_k}, Z_{obs_k}$ are observed indicators of the corresponding underlying latent variables. The matrices D_{yk}, D_{xk}, D_{zk} are composed of known factor loadings that link the latent variables to observed indicators, and the measurement errors e_{yk}, e_{xk}, e_{zk} are assumed multivariate normal in distribution with zero means and covariance structure that depend heavily on the study design, as illustrated in the example below.

TRANSFORMED STRUCTURAL MODEL

Those familiar with hierarchical models will see that the measurement model of equation (5) is equivalent to a level-1 model where the observed outcomes vary as a function of X_k, Y_k, and Z_k, conceived here as random coefficients. The level-2 model would describe the joint distribution of X_k, Y_k, and Z_k or, equivalently, the conditional distribution of Y_k and Z_k given the exogenous X_k.[3] Combining equations (2) and (3),

$$\begin{pmatrix} Y_k \\ Z_k \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{B}_{yx} X_k \\ \mathbf{B}_{zx} X_k \end{pmatrix}, \begin{pmatrix} \mathbf{T}_{yy} & \mathbf{T}_{yz} \\ \mathbf{T}_{zy} & \mathbf{T}_{zz} \end{pmatrix}\right], \tag{6}$$

implying

$$\begin{aligned} \Gamma_{yz.x} &= \mathbf{T}_{yz} \mathbf{T}_{zz}^{-1} \\ \mathbf{B}_{yx.z} &= \mathbf{B}_{yx} - \Gamma_{yz.x} \mathbf{B}_{zx} \\ \Delta &= \Gamma_{yz.x} \mathbf{B}_{zx} \\ \Sigma &= \mathbf{T}_{yy} - \mathbf{T}_{yz} \mathbf{T}_{zz}^{-1} \mathbf{T}_{zy}. \end{aligned} \tag{7}$$

Equation (7) expresses the parameters of the simultaneous equation model (equations (1) and (3)) as one-to-one transformations of the parameters of the joint distribution of *Y* and *Z* (given *X*). The parameters of this joint distribution are conveniently estimated within the framework of ML. ML estimates of the parameters of the simultaneous equation model follow.

ESTIMATION

The task now is to estimate the distribution of the latent variables given by equation (6) and then to employ the transformations of equation (7) in order to obtain the desired inferences with regard to the direct effects, $\Gamma_{yx.z}$, and the indirect effects, Δ. Let β be the vector of elements of ($\mathbf{B}_{yx}$, $\mathbf{B}_{zx}$), and let τ be the vector of unique elements of ($\mathbf{T}_{yy}$, $\mathbf{T}_{yz}$, $\mathbf{T}_{zz}$). Now-standard software provides ML estimates, for example via the EM algorithm (Dempster, Rubin, and Tsutakawa 1981; Bryk and Raudenbush 1992, chap. 10) or the Fisher scoring algorithm (Longford 1987). We seek ML point estimates and intervals for the transformations of (β, τ) given by equation (7). The ML point estimates are, of course, the corresponding transformations of the estimates ($\hat{\beta}$, $\hat{\tau}$), based on the invariance properties of ML estimation. Variances of these estimators are readily derived from first-order (or higher order) Taylor series expansions.[4]

To implement this simple idea, we express all estimands as vectors:

$$
\begin{aligned}
\beta_{yx} &= vec\left(\mathbf{B}_{yx}^{T}\right)\\
\gamma_{yz.x} &= vec\left(\Gamma_{yz.x}^{T}\right) = vec\left[\left(\mathbf{T}_{yz}\mathbf{T}_{zz}^{-1}\right)^{T}\right]\\
\beta_{yx.z} &= vec\left(\mathbf{B}_{yx.z}^{T}\right) = vec\left[\left(\mathbf{B}_{yx} - \Gamma_{yz.x}\mathbf{B}_{zx}\right)^{T}\right]\\
\delta &= vec\Delta^{T} = vec\left[\left(\mathbf{B}_{yx} - \mathbf{B}_{yx.z}\right)^{T}\right] = vec\left[\left(\Gamma_{yx\cdot z}\mathbf{B}_{zx}\right)^{T}\right].
\end{aligned}
\tag{8}
$$

The vectors on the left-hand side of equation (8) are all simple one-to-one transformations of parameters routinely estimated in hierarchical models. Appendix A derives the asymptotic variances of the needed estimates. These are all functions of the Fisher information routinely computed in conjunction with the hierarchical model.

ILLUSTRATIVE EXAMPLE

The question motivating the illustrative analysis is whether the relationship between the social composition of urban neighborhoods and levels of violence in those neighborhoods is explained (mediated) by measurable characteristics of neighborhood social organization. Data for the analysis were collected during 1995 under the auspices of the Project on Human Development in Chicago Neighborhoods (PHDCN). The results reported here are intended to illustrate the methodology rather than to provide conclusive substantive evidence. A more fine-grained analysis making stronger claims about the importance of neighborhood social organization appears in Sampson, Raudenbush, and Earls (1997).

SAMPLING DESIGN AND DATA

The design of PHDCN's community survey involved, first, the assignment of each of Chicago's 847 census tracts to 342 geographically contiguous neighborhood clusters (NCs) constituted to be internally homogeneous with respect to resident socioeconomic status, ethnic mix, and housing density while preserving physical boundaries of neighborhoods. Within each NC, a probability sample of households was selected, and within each household, a capable adult interviewed with regard to conditions, events, and relationships within the local area that the resident defined as "the neighborhood." Responses to related questions were combined into scales representing sociologically important aspects of neighborhood social organization. These scales can also be aggregated to the NC level to construct measures of NC social organization.

Our current interest focuses on "social control," a five-item scale based on extant theory and extensive pretesting. Residents were asked about the likelihood that their neighbors could be counted on to intervene in various ways if (1) children were skipping school and hanging on a street corner, (2) children were spray painting graffiti on a local building, (3) children were showing disrespect to an adult, (4) a fight broke out in front of their house, and (5) the fire station closest to home was threatened with budget cuts. Responses were on a 5-point Likert-type scale. Most respondents answered all five questions; for those respondents, the scale score was the average of the five responses. However, anyone responding to at least one item provided data for the analysis; a person-specific standard error of measurement was calculated based on a simple linear item-response model that took into account the number and "difficulty" of the items to which each resident responded (see Appendix B for details).

Respondents were also asked five questions with regard to the occurrence of incidents of violence in the neighborhood. Specifically, they were asked how often each of the following occurred in the neighborhood during the past 6 months: (1) a fight in which a weapon was used, (2) a violent argument between neighbors, (3) gang fights, (4) a sexual assault or rape, and (5) a robbery or mugging. Scale construction for perceived violence mirrored that for social control.

Reasonably complete data are available from 7,726 persons residing in 342 NCs, on average about 23 per NC.[5] In effect, the aim of the community survey is to use neighborhood residents as *informants* about neighborhood social organization of each NC; that is, the key units being assessed are the NCs. Because the number of "raters" per NC varies, so does the reliability of measurement. The resulting measurement and analytic issues in studies of this sort are described in detail by Raudenbush et al. (1991), although in their examples teachers are informants about the social organization of their schools.

NC social composition was measured independently from the 1990 decennial census. These include poverty concentration (percent below the poverty line), ethnic isolation (percent African American), and percent foreign born, each of which is believed to be positively associated with perceived violence within NCs. Social composition has no hypothesized direct link to violence; rather, we expect that neighborhoods with disadvantaged social composition are less likely than more advantaged neighborhoods to be effectively organized to monitor, supervise, or otherwise control the social behavior of young people living in the neighborhoods, and that this lack of social control will help explain the statistical link between social composition and violence. For a theoretical explication of this argument, see Sampson and Groves (1989) and Sampson and Lauritsen (1994).

TABLE 1: Description of the Sample

Variable	n	*Mean*	SD	*Minimum*	*Maximum*
Person-level data					
Gender[a]	7,726	0.59	0.49	0.00	1.00
Age	7,726	42.59	16.73	17.00	100.00
Socioeconomic status[b]	7,726	0.00	1.32	–4.08	4.33
Neighborhood cluster–level data					
Poverty concentration	342	20.43	17.31	0.23	88.18
Ethnic isolation	342	41.21	43.67	0.00	99.81
Percent foreign born	342	16.54	15.63	0.00	64.62
Violence[c]	342	1.88	0.41	1.13	3.17
Social control	342	3.49	0.40	2.38	4.63

a. Coded as 1 = female, 0 = male.
b. Socioeconomic status is the first principal component of household income, respondent education, and respondent occupation.
c. Violence and social control are the neighborhood cluster means of the observed scores for persons in that cluster.

Table 1 describes the sample. We see that the mean percent African American across the 342 neighborhood clusters is 41.2 with a large standard deviation. The mean percent below the poverty line (about 20) and percent foreign born (about 16.5) reflect Chicago's diverse population.

It is also important to control for social selection; that is, within a neighborhood, survey responses may be shaped by the socioeconomic status (SES), age, gender, and so on of the respondents, and this link between social-demographic background and responses as it occurs *within* neighborhoods should be statistically controlled in an analysis that seeks to understand variation and covariation in outcomes *between* neighborhoods. Otherwise, the composition of the sample with respect to age, SES, gender, and so on will bias the measure of NC characteristics. Table 1 describes the person-level covariates used in the illustrative analysis: age, gender (coded 1 for females, 0 for males) and SES (the first principal component of the respondent's years of education, occupation status, and income).

MODEL

An analytic model is needed that will (1) control for the varying measurement error of survey-based measures of social control and perceived violence; (2) control for personal characteristics related to responses within neighborhoods; and (3) appropriately account for the clustered nature of the sample, in which item responses are nested within persons who are themselves nested within NCs. A three-level hierarchical model is appropriate to this task. Raudenbush et al. (1991) provide a detailed presentation of the estimation theory. The model is specified in three stages to clarify key assumptions about measurement error and clustering.

Level-1 Model

At the first level, we model the error with which perceived violence and social control are measured:

$$\begin{aligned} R_{ijk} &= D_{1ijk}\left(Y_{jk} + \varepsilon_{jk}\right) + D_{2ijk}\left(Z_{jk} + \nu_{jk}\right) \\ \varepsilon_{jk} &\sim N\left(0, \sigma^2_{1jk}\right) \\ \nu_{jk} &\sim N\left(0, \sigma^2_{2jk}\right). \end{aligned} \tag{9}$$

Equation (9) may be viewed as a classical measurement model in which R_{ijk} is a fallible measure of latent variable i for person j living in neighborhood k. In this example, there are two latent variables: Y_{jk}, the "true" value of perceived violence in neighborhood k as reported by person j, and Z_{jk}, the "true" level of neighborhood social control in neighborhood k as perceived by person j.[6] The predictor D_{1ijk} is an indicator variable taking a value of 1 if R_{ijk} is a measure of perceived violence and 0 if not; similarly, D_{2ijk} takes a value of 1 if R_{ijk} measures social control and 0 if not (i = 1,2; j = 1, . . ., J_k; k = 1, . . ., K). This formulation allows for the utilization of all available data in the analysis, avoiding, for example, the listwise deletion of persons giving data on social control but not perceived violence. The error variance with which R_{ijk} measures the intended latent variable is either σ^2_{1jk} or σ^2_{2jk}. Several approaches are available for modeling these measurement error variances within the framework of the hierarchical model. We estimated these variances from a separate item response analysis. Alternatively, they can be estimated simultaneously with all other model parameters. Both approaches are described in detail in Appendix B.

To clarify exactly how the level-1 model works, consider a respondent with available data on both perceived violence and social control. The first observed measure, R_{1jk}, would be the fallible measure of perceived violence, and the second measure, R_{2jk}, would be the fallible measure of social control. For such a respondent, the level-1 model would consist of the pair of equations

$$\begin{aligned} R_{1jk} &= (1)(Y_{jk} + \varepsilon_{jk}) + (0)(Z_{jk} + \upsilon_{jk}) \\ &= Y_{jk} + \varepsilon_{jk} \\ \mathrm{R}_{2jk} &= (0)(Y_{jk} + \varepsilon_{jk}) + (1)(Z_{jk} + \upsilon_{jk}) \\ &= Z_{jk} + \upsilon_{jk}. \end{aligned} \tag{10}$$

Suppose, however, that person *jk* supplied data on perceived violence but not social control. Then, the level-1 model would consist of a single equation

$$\begin{aligned} R_{1jk} &= (1)(Y_{jk} + \varepsilon_{jk}) + (0)(Z_{jk} + \upsilon_{jk}) \\ &= Y_{jk} + \varepsilon_{jk}. \end{aligned} \tag{11}$$

Similarly, if person *jk* provided only the social control data, the first and only fallible measure, R_{1jk} would be a measure of Z_{jk}, and the level-1 model would be

$$\begin{aligned} R_{1jk} &= (0)(Y_{jk} + \varepsilon_{jk}) + (1)(Z_{jk} + \upsilon_{jk}) \\ &= Z_{jk} + \upsilon_{jk}. \end{aligned} \tag{12}$$

An important feature of the model is that, even for respondents who supply no data on perceived violence or social control, the "true scores" Y_{jk} and Z_{jk}, in principle, exist. We seek inferences about the distribution of the complete latent data given whatever observed data are available.

Level-2 Model

The second level describes variation in the two latent variables within neighborhoods:

$$
\begin{aligned}
Y_{jk} &= Y_k + \pi_{y1k}(age)_{jk} + \pi_{y2k}(gender)_{jk} + \pi_{y3k}(SES)_{jk} + r_{yjk} \\
Z_{jk} &= Z_k + \pi_{z1k}(age)_{jk} + \pi_{z2k}(gender)_{jk} + \pi_{z3k}(SES)_{jk} + r_{zjk} \\
&\begin{pmatrix} r_{yjk} \\ r_{zjk} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{pmatrix}\right].
\end{aligned}
\tag{13}
$$

Thus, within neighborhoods, latent responses are viewed as possibly dependent on age (in years, grand mean centered), gender (1 = female, 0 = male), and SES (the first principal component of household income, respondent occupational prestige, and respondent years of education). Of central interest in this analysis are Y_k and Z_k, the “true” neighborhood means on perceived violence and social control, adjusted for the possible within-neighborhood response biases linked to age, gender, and SES.[7] The random effects r_{yjk} and r_{zjk}, having covariance Ω_{yz}, capture the dependence among multiple responses within persons, conditional on Y_k, Z_k, and the πs in equation (13).

Level-3 Model

The third and final level of the model describes the variation across neighborhoods of adjusted neighborhood mean perceived violence and social control:

$$
\begin{aligned}
Y_k &= \beta_{y0} + \beta_{y1}(pov\ con)_k + \beta_{y2}(ethnic\ iso)_k \\
&\quad + \beta_{y3}(\%\,for\ born)_k + u_{yk} \\
Z_k &= \beta_{z0} + \beta_{z1}(pov\ con)_k + \beta_{z2}(ethnic\ iso)_k \\
&\quad + \beta_{z3}(\%\,for\ born)_k + u_{zk} \\
&\begin{pmatrix} u_{yk} \\ u_{zk} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} T_{yy} & T_{yz} \\ T_{zy} & T_{zz} \end{pmatrix}\right].
\end{aligned}
\tag{14}
$$

Thus, the adjusted mean perceived violence, Y_k, and social control, Z_k, vary across NCs as a function of poverty concentration, ethnic isolation, and percent foreign born plus a pair of random effects (u_{yk}, u_{zk}) assumed to be bivariate normal in distribution. These random effects capture the dependence between persons living in the same neighborhood.

Equation (14) is recognizable as a special case of equation (6) with

$$X_k = [1 \quad (pov\ con)_k \quad (ethnic\ iso)_k \quad (\%\ for\ born)_k\ ]^T. \tag{15}$$

We face choices with regard to the level-3 specification of regression coefficients of the level-2 model. For parsimony, all within-neighborhood regression coefficients other than the adjusted mean (i.e., all π_{yfk} and π_{zfk} for $f = 1,2,3$) are constrained to be constants. Formally, we have

$$\pi_{yfk} = \beta_{yf0} \quad \pi_{zfk} = \beta_{zf0} \quad f = 1,2,3. \tag{16}$$

In fact, the model and estimation procedure can readily incorporate the random variation of such coefficients. Suppose, for example, that the association between age and perceived violence varies randomly over neighborhoods. The model for perceived violence might then be

$$\pi_{y1k} = \beta_{y10} + u_{y1k}, \tag{17}$$

where u_{y1k} is the random effect of neighborhood k. In our case, no theory is available in this application to suspect or interpret such variation; we therefore constrain these coefficients in the interest of parsimony. Instances in which such variation may be of interest are briefly considered in the summary and discussion.

In sum, the level-1 and level-2 models describe the sources of measurement error whereas the level-3 model describes the joint distribution of the latent variables. Specifically, the level-1 model (equation (9)) describes how item responses are aggregated to measure respondent perceptions and the level-2 model (equation (13)) describes how respondent perceptions are aggregated to indicate neighborhood social organization. The level-3 model (equation (14)) is of central theoretical interest because it characterizes the associations between neighborhood-level constructs. In other applications, it may be that latent variables of key theoretical interest are at the person level or at both person and neighborhood levels. In such cases, the level-1 model might be the measurement model whereas the level-2 and level-3 models would describe the two-level structure of the key constructs.

TABLE 2: Perceived Violence and Social Control as a Function of Neighborhood Social Composition

	Perceived Violence			*Social Control*		
	Estimated Coefficient	*Standard Error*	*Ratio*	*Estimated Coefficient*	*Standard Error*	*Ratio*
Intercept						
Level-1 predictors						
Gender	0.880	1.734	0.51	0.481	2.051	0.23
Age	–0.521	0.053	–9.80	0.306	0.062	4.95
Socioeconomic status	1.024	0.725	1.41	2.863	0.849	3.73
Level-2 predictors						
Poverty concentration	1.285	0.105	12.21	–0.912	0.108	–8.47
Ethnic isolation	0.352	0.060	5.87	–0.462	0.062	–7.49
Percent foreign born	0.636	0.147	4.34	–1.246	0.151	–8.24
Within–neighborhood cluster covariance components						
Variances	3925	86		5829	124	
Covariance	–1601	76				
Between–neighborhood cluster covariance components						
Variances	443	54		375	56	
Covariance	–205	42				

NOTE: In this and other tables, social control and perceived violence scales were multiplied by 100 to ensure printing of a reasonable number of significant digits without resorting to scientific notation.

RESULTS

The three-level model was estimated via ML as described in detail by Raudenbush et al. (1991). At convergence, the asymptotic variance matrix of the estimated coefficients (β) and the estimated variance-covariance elements ω in (Ω_{yy}, Ω_{yz}, Ω_{zz}) and τ in ($\mathbf{T}_{yy}$, $\mathbf{T}_{yz}$, $\mathbf{T}_{zz}$) are routinely computed using software for three-level models. Results are provided in Table 2.

Controlling Response Bias

As Table 2 indicates, age (but not gender or SES) is linked to resident responses to questions about violence. Older residents report

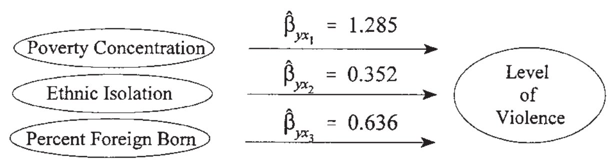

Figure 3: Total Effect of Exogenous Predictors (poverty concentration, ethnic isolation, percent foreign born) on Levels of Violence

lower levels of violence than do younger residents within the same NC ($\hat{\beta}_{y10}$ = –0.521, t = –9.80). Age and SES (but not gender) are associated with responses to questions about social control. Age and SES are positively related to the perceptions that neighbors are willing to exercise social control ($\hat{\beta}_{z10}$ = 0.306, t = 4.95; $\hat{\beta}_{z30}$ = 2.86, t = 3.73). Thus, there is some evidence that demographic differences between persons living in the same neighborhood are associated with their perceptions about that neighborhood. This source of bias is controlled in examining between-neighborhood associations.

Total Effect of X *on* Y

As hypothesized, neighborhood poverty concentration ($\hat{\beta}_{y1}$ = 1.285, t = 12.21), ethnic isolation ($\hat{\beta}_{y2}$ = 0.352, t = 5.87), and percent foreign born ($\hat{\beta}_{y3}$ = 0.636, t = 4.34) are positively related to perceived violence. These associations are depicted graphically in Figure 3.

Association Between X *and* Z

Neighborhood poverty concentration, ethnic isolation, and percent foreign born are negatively linked to social control ($\hat{\beta}_{z1}$ = –0.912, t = –8.47; $\hat{\beta}_{z2}$ = –0.462, t = –7.49; and $\hat{\beta}_{z3}$ = –1.246, t = –8.24; see Figure 4).

Transformation of Hierarchical Model Estimates to Estimate Direct and Indirect Effects

We now transform equation (14), which specifies the distribution of Y_k, $Z_k|X_k$ to estimate the distribution of $Y_k|Z_k$, X_k. The new model is

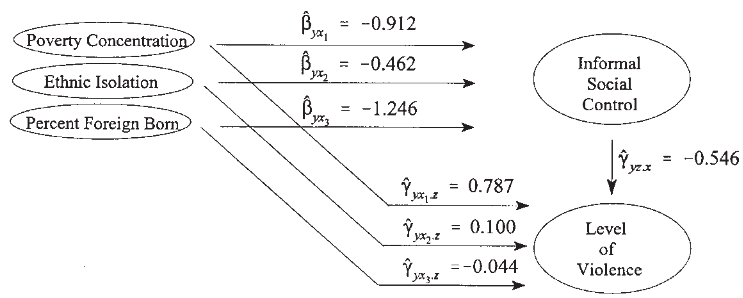

Figure 4: Direct and Indirect Effects of Exogenous Predictors on Level of Violence With Informal Social Control as a Mediator

$$\begin{aligned} Y_k &= \beta_{y0.z} + \beta_{y1.z}(pov\ con)_k + \beta_{y2.z}(ethnic\ iso)_k \\ &+ \beta_{y3.z}(\%for\ born)_k + \gamma_{yz.x} Z_k + u_k, \quad u_k \sim N\left(0, \sigma^2\right). \end{aligned} \tag{18}$$

Note that the latent social control measure, Z_k, which had been one of the two outcome variables (equation (14)), has now become a latent predictor variable on the right-hand side of equation (18). Of interest are the association between that variable and perceived violence and the direct effects of neighborhood social composition (X) on Y, and the indirect effects of neighborhood social composition on Y as mediated by Z, social control.

Equation (18) produces a special case of equation (8) with

$$\begin{aligned} \beta_{yx.z} &= \left(\beta_{y0.z}, \beta_{y1.z}, \beta_{y2.z}, \beta_{y3.z}\right)^T \\ \gamma_{yz.x} &= \gamma_{yz.x} \\ \delta &= \left(\gamma_{yz.x}\beta_{z0}, \gamma_{yz.x}\beta_{z1}, \gamma_{yz.x}\beta_{z2}, \gamma_{yz.x}\beta_{z3}\right)^T \\ \Sigma &= \sigma^2 = T_{yy} - T_{yz}^2 / T_{zz}. \end{aligned} \tag{19}$$

Association Between Z and Y Given X

As Table 3 indicates, there is strong evidence of a negative association between social control and perceived violence ($\hat{\gamma}_{yz.x} = -0.546$, $t = -5.55$) net the contributions of social composition (poverty concentration, ethnic isolation, and percent foreign born) (see also Figure 4).

TABLE 3: Perceived Violence as a Function of Social Control and Neighborhood Social Composition

	Estimated Coefficient	*Standard Error*	*Ratio*
Intercept	375.00	39.07	9.60
Level-2 predictors			
Poverty concentration	0.787	0.133	5.90
Ethnic isolation	0.100	0.072	1.39
Percent foreign born	–0.044	0.184	–0.24
Social control	–0.546	0.098	–5.55
Between–neighborhood cluster variance	331		

Direct Effect of X on Y

Adjustment for social control reduces the contributions of ethnic isolation and percent foreign born to nonsignificance ($\hat{\beta}_{y2.z}$ = 0.100, t = 1.39; $\hat{\beta}_{y3.z}$ = –0.044, t = –0.24). Poverty concentration remains significantly positively related to perceived violence ($\hat{\beta}_{y1.z}$ = 0.787, t = 5.90), but its coefficient appears to be considerably smaller than prior to the adjustment for social control; compare coefficients for poverty concentration in Table 2 ($\hat{\beta}_{y01}$ = 1.285) and Table 3 ($\hat{\beta}_{y01.z}$ = 0.787) (see Figure 4).

Indirect Effects

Table 4 lists the total effects, the direct effects, and the indirect effects. It also tabulates the standard relevant errors. We see substantial indirect effects for each aspect of social composition. Thus, in each case, social composition is linked to violence through its association with social control. The magnitude of each indirect effect is far larger than its standard error, providing evidence of statistically significant mediating effects of social control.

Comparing Coefficients Between Models

The indirect effects in Table 4 can also be interpreted as the difference between the total effect (association between X and Y) and the direct effect (association between X and Y controlling Z). This interpre-

TABLE 4: Decomposition of Total Effects Into Direct and Indirect Components

	Total	*Direct*	*Indirect*
Poverty concentration	1.285	0.787	0.498
	(0.105)	(0.133)	(0.107)
Ethnic isolation	0.352	0.100	0.253
	(0.060)	(0.072)	(0.056)
Percent foreign born	0.636	–0.044	0.680
	(0.147)	(0.184)	(0.148)

NOTE: Standard errors in parentheses.

tation is discussed in Clogg et al. (1995) and Allison (1995). In each case, the large size of the difference relative to its estimated standard error implies that the differences between the relevant coefficients are significantly greater than zero. The standard errors can be used to compute confidence intervals for such differences.

Summary

Based on the results of Tables 2-4, we can conclude that (1) the three social composition indicators are associated positively as hypothesized with perceived violence (total effects); (2) given social composition, neighborhood social control is negatively related to perceived violence; (3) adjusting for social control, we find a statistically significant direct effect of concentrated poverty on violence but no significant direct effects between either ethnic isolation or percent foreign born and violence; (4) links between *X* (poverty concentration, ethnic isolation, and percent foreign born) and perceived violence are partially explained or mediated by neighborhood social control (as indicated by statistically significant indirect effects). This system of simultaneous equations is depicted in Figure 4.

Importance of Controlling for Errors of Measurement of Neighborhood Social Control

To assess the impact of the latent variable model on statistical inference, we computed a multilevel analysis of perceived violence with exactly the same level-1 and level-2 predictors except that manifest social control rather than latent social control was included as a level-2

TABLE 5: Consequences of Modeling Measurement Error in Social Control

	Latent Social Control as Predictor			*Manifest Social Control as Predictor*		
	Estimated Coefficient	*Standard Error*	*Ratio*	*Estimated Coefficient*	*Standard Error*	*Ratio*
Level-2 predictors						
Poverty concentration	0.787	0.133	5.90	0.877	0.108	8.13
Ethnic isolation	0.100	0.072	1.39	0.157	0.059	2.64
Percent foreign born	–0.044	0.184	–0.24	0.097	0.148	0.66
Social control	–0.546	0.098	–5.55	–0.412	0.048	–8.63
Between–neighborhood cluster variance	331			327		

predictor. The manifest version of social control was computed as the neighborhood cluster mean response to the social control scale of sample responses within each NC (descriptive statistics in Table 1). The reliability of this measure at the NC level depends heavily on the sample size of informants per NC. As sample sizes range from 20 to 50, the reliability ranges from .70 to .86. Whereas the latent variable analysis takes this varying reliability into account, the analysis based on the manifest variable does not. The results (Table 5) show, as might be expected, that the estimate of the social control coefficient is smaller when the manifest social control is used than when the latent variable model is estimated (–.412 as compared to –.546). Note also that the estimated standard error of the coefficient is substantially larger in the case of the latent variable analysis, reflecting the additional uncertainty with regard to the error of measurement of social control. Also, the adjustments to the contributions of ethnic isolation and percent foreign born are less severe when manifest social control rather than latent social control is included in the model. In fact, ethnic isolation retains a significant direct effect on violence in the full model only when the manifest indicator of social control is used.

In sum, if one ignores measurement error of social control, one's belief about the importance of social control relative to that of social composition is diminished, the confidence interval for the social control coefficient is shortened, and social control appears less important as a mediator of the association between social composition and perceived violence. That is, the direct effects of social composition on violence are exaggerated.

SUMMARY AND DISCUSSION

In a two-level hierarchical model, the first level describes the distribution of an observed outcome given a set of random coefficients defined on level-2 units. The second level of the model describes the distribution of those random coefficients over the level-2 units. This article has adapted the hierarchical model to describe direct and indirect effects involving latent variables. A measurement model describes the distribution of the observed data given the latent variables, which play the role of random coefficients. A structural model describes the joint distribution of the latent variables. Total, direct, and indirect effects are one-to-one transformations of the parameters of that joint distribution and are thus estimable via ML using standard algorithms for hierarchical models. What are the strengths and limitations of this setup in the study of mediated effects among latent variables?

The strengths of the approach are several. First, the approach readily handles the kind of unbalanced data generated by multilevel sampling designs in large-scale surveys. In the example, survey respondents were nested within urban neighborhoods, and the sample size varied quite widely over neighborhoods. The general approach can readily incorporate this and more complex multilevel designs.

Second, the approach naturally incorporates all available information (see equations (10)-(12) and associated discussion). Often, survey respondents will not be available at all interviews, and some may refuse to answer certain questions. Valid inference depends on the assumption that the missingness is ignorable (cf. Little and Schenker 1995), but the robustness of the results to nonignorable missingness is generally greatest when all available information is used in the analysis (Schafer 1997). In other cases, it is too expensive to collect all data on every respondent, so a sampling scheme is used to determine which respondents will provide which data. This is true, for example, in the National Adult Literacy Survey and the National Assessment of

Educational Progress of the National Center for Educational Statistics. In these cases, the missingness is ignorable if all available information is appropriately used in the analysis.

Third, the approach naturally incorporates random coefficients, but only of those predictors measured without error.[8] Random coefficients were not of interest in the illustrative example, but in other studies, for example, longitudinal studies of varying growth rates (Bryk and Raudenbush 1992, chap. 6), will be of central interest. The growth rate is typically the random coefficient associated with age. Age is typically measured accurately, but the predictors of growth may be measured fallibly. The approach described here can handle fallible predictors of randomly varying growth rates.

Fourth, all parameters are estimated simultaneously via ML. This is a major strength of common approaches for estimating structural equation models. Thus, the method is not vulnerable to Allison's (1995) criticism of approaches that base inferences about the total effect of *X* on *Y* on naive models that assume, for example, uncorrelated and homoscedastic residuals (see equation (3), which gives the assumed structure of the residuals in estimating the total effects of *X* on *Y*).

Finally, the model provides a straightforward way to compare coefficients between models (Clogg et al. 1995; Allison 1995). The aim in such analyses is to compare the association between *X* and *Y* with and without adjustment for *Z*.

Limitations are also apparent. First, inferences may not be robust to the nonnormality of *Y* and *Z* given *X*. In particular, estimation of $\Gamma_{yz.x}$ (the association between *Z* and *Y* given *X*) is derived from the ML estimate of the variance-covariance matrix of *Y* and *Z* given *X* (see equation (7)). This ML estimate assumes multivariate normality, and covariance component estimates are generally less robust to violations of this assumption than are the usual regression coefficients. In our example, the central limit theorem works in favor of normality because multiple item responses aggregated across multiple informants provide information about *Y* and *Z*. Nevertheless, in all cases, residuals and *Q-Q* plots should be checked carefully (see Bryk and Raudenbush 1992, chap. 9). This problem arises generally in covariance structure analyses.

Second, the estimability of the model and stability of inferences depend strongly on the variance-covariance structure of *Z* (note from equation (7) that the ML estimate of $\mathbf{T}_{zz}$ must be inverted to decompose the total effect of *X* on *Y* into direct and indirect components). The following conditions, alone but especially in combination, are likely to cause trouble: having many *Z* variables, low reliabilities in estimating *Z*, small variances of *Z*, and collinearity among them. In our example, there was a single *Z* variable estimated with reasonable reliability in most NCs. In the absence of more research, we recommend parsimonious specification of mediators *Z*, careful checking of reliabilities, and a close look at the ML estimate of $\mathbf{T}_{zz}$ and its information matrix.

Third, all inferences are based on large-sample theory. Our example used 342 NCs. With parsimonious model specification, we feel confident about the resulting large-sample inferences. Given the novelty of multilevel latent variable modeling, little is known about the behavior of the estimates in small or moderate-sized samples. However, we can confidently anticipate that the parsimony of model specification operates in conjunction with large sample size to support sound inference. More research is needed with regard to small-sample properties of the estimators and possibly improved approximations for the standard errors.

Fourth, most structural modelers would consider the class of measurement models currently available in our approach to be limited. As articulated in this article, the approach requires known factor loadings and yields latent variables with varying variances. Such an approach requires that all items loading on a given factor be measured on the same scale, whereas the factor variances and covariances are allowed to vary freely. This differs from a common approach in latent variable models whereby the factor loadings are estimated from the data and the variances of the latent variables are constrained to unity. The appeal of using known factor loadings (typically set to 1.0) is a potential robustness across studies. However, it would be useful to allow a wider class of measurement models. The task essentially involves allowing a wider class of covariance structures (cf. Jennrich and Schluchter 1986). This task can readily be tackled within the framework of our model, although we are not aware of any currently available software that can incorporate this wider class of structures for these models.

The work reported here is part of a broader trend toward the formulation of statistical models that are realistically flexible and complex, for example, in accommodating multilevel designs, simultaneous equations, measurement error, and missing data. Enthusiasm over new modeling possibilities must, however, be accompanied by cautions with regard to possible pitfalls in applying these more complex models. Such pitfalls include the need for new assumptions and the associated possible lack of robustness and a loss of precision that occurs when the available data are used to estimate ever more parameters. The best antidotes appear to be a commitment to parsimony in model specification and a commitment to careful data analysis, including a check of assumptions and an assessment of the sensitivity of results to plausible alternative specifications.

APPENDIX A
Computation of Standard Errors

This appendix derives asymptotic variances of estimates of the simultaneous equation model parameters (left-hand side of equation (7)). These estimates are one-to-one transformations of parameters routinely estimated by hierarchical models (right-hand side of equation (7)), and our strategy in computing the needed variances is simply to compute the relevant transformations of the information matrix computed in the context of the hierarchical model.

Note from equation (7) that whereas $\gamma_{yz.x}$ depends only on τ, $\beta_{yx.z}$ and δ depend on both β and τ, and that the ML estimators of β and τ are independent. First-order expansions and variance expressions follow:

$$\hat{\gamma}_{yz.x} \approx \gamma_{yz.x} + \frac{\partial \gamma_{yz.x}}{\partial \tau^{\mathrm{T}}}(\hat{\tau} - \tau)$$
$$Var(\hat{\tau}_{yz.x}) \approx \frac{\partial \gamma_{yz.x}}{\partial \tau^{T}} Var(\hat{\tau}) \left[\frac{\partial \gamma_{yz.x}}{\partial \tau^{T}}\right]^{T} \quad (20)$$

$$\begin{aligned}\hat{\beta}_{yx.z} &\approx \beta_{yx.z} + \left(\frac{\partial\beta_{yx.z}}{\partial\beta^T}\ \frac{\partial\beta_{yx.z}}{\partial\tau^T}\right)\begin{bmatrix}\hat{\beta}-\beta\\ \hat{\tau}-\tau\end{bmatrix}\\ Var\left(\hat{\beta}_{yx.z}\right) &\approx \frac{\partial\beta_{yx.z}}{\partial\beta^T} Var\left(\hat{\beta}\right)\left[\frac{\partial\beta_{yx.z}}{\partial\beta^T}\right]^T\\ &\quad + \frac{\partial\beta_{yx.z}}{\partial\tau^T} Var(\hat{\tau})\left[\frac{\partial\beta_{yx.z}}{\partial\tau^T}\right]^T\end{aligned} \tag{21}$$

and

$$\begin{aligned}\hat{\delta} &\approx \delta + \left(\frac{\partial\delta}{\partial\beta^T}\ \frac{\partial\delta}{\partial\tau^T}\right)\begin{bmatrix}\hat{\beta}-\beta\\ \hat{\tau}-\tau\end{bmatrix}\\ Var\left(\hat{\delta}\right) &\approx \frac{\partial\delta}{\partial\beta^T} Var\left(\hat{\beta}\right)\left[\frac{\partial\delta}{\partial\beta^T}\right]^T + \frac{\partial\delta}{\partial\tau^T} Var(\hat{\tau})\left[\frac{\partial\delta}{\partial\tau^T}\right]^T.\end{aligned} \tag{22}$$

In equations (20)-(22), all partial derivatives are evaluated at the ML estimators of their components. The required derivative matrices are readily evaluated column by column. Let τ_{rs} denote element *rs* of τ. Then, we have

$$\begin{aligned}\frac{\partial\gamma_{yz.x}}{\partial\tau_{rs}} &= vec\left(\frac{\partial\Gamma^T_{yz.x}}{\partial\tau_{rs}}\right)\\ &= vec\left[\left(C_{yzrs} - \Gamma_{yz.x}C_{zzrs}\right)\mathbf{T}_{zz}^{-1}\right]^T,\end{aligned} \tag{23}$$

where

$$C_{yzrs} = \frac{\partial\mathbf{T}_{yz}}{\partial\tau_{rs}}, \qquad C_{zzrs} = \frac{\partial\mathbf{T}_{zz}}{\partial\tau_{rs}}. \tag{24}$$

An element of C_{yzrs} (or C_{zzrs}) is equal to unity when the corresponding element of $\mathbf{T}_{yz}$ (or $\mathbf{T}_{zz}$) is equal to τ_{rs}; other elements are equal to zero. Similarly,

$$\frac{\partial\beta_{yx.z}}{\partial\tau_{rs}} = -\frac{\partial\delta}{\partial\tau_{rs}} = -vec\left[\left(\frac{\partial\Gamma_{yz.x}}{\partial\tau_{rs}}\mathbf{B}_{zx}\right)^T\right]. \tag{25}$$

Let β_{hl} denote element *hl* of β. Then,

$$\frac{\partial \beta_{yx.z}}{\partial \beta_{hl}} = vec\left(\frac{\partial \mathbf{B}_{yx}}{\partial \beta_{hl}} - \Gamma_{yz.x}\frac{\partial \mathbf{B}_{zx}}{\partial \beta_{hl}}\right)^T \tag{26}$$

and

$$\frac{\partial \delta}{\partial \beta_{hl}} = vec\left(\Gamma_{yz.x}\frac{\partial \mathbf{B}_{zx}}{\partial \beta_{hl}}\right)^T. \tag{27}$$

APPENDIX B
Measurement Error Models

Several approaches are available to estimate measurement error variances associated with the level-1 model (see equation (9)). We estimated these variances in a separate item analysis. The outcomes at level 1 were the 10 possible item responses, 5 to questions about social control and 5 to questions about perceived violence. The resulting level-1 model was

$$\begin{aligned} S_{ijk} &= D_{1ijk}\left(Y_{jk} + \sum_{p=1}^{4} E_{pijk}\alpha_{pjk} + \varepsilon_{jk}\right) \\ &+ D_{2ijk}\left(Z_{jk} + \sum_{q=1}^{4} E_{qijk}\alpha_{qjk} + \upsilon_{jk}\right). \end{aligned} \tag{28}$$

As in equation (9), the dummy variables D_{1ijk} and D_{2ijk} indicate, respectively, whether item response S_{ijk} is associated with perceived violence or social control. There are eight other dummies, E_{pijk} ($p = 1, \ldots, 4$) and E_{qijk} ($p = 1, \ldots, 4$), four for items measuring perceived violence and four for items measuring social control. The coefficients α_{pjk} and α_{qjk} are "item difficulties," whereas Y_{jk} and Z_{jk} are "true scores" or latent variables. The benefit of this approach is that for respondents who fail to respond to all items, these true scores are adjusted for the difficulty of the items to which they did respond. The level-2 and level-3 models include no covariates; the item difficulties are constrained to be fixed coefficients. The model produces an estimated true score and a standard error for every respondent who answered at least one of the questions within a scale. These estimated true scores and their associated squared standard errors were used in equation (9).

A closely related alternative approach is to use a nonlinear item response model, for example a one-parameter logit model for binary or ordinal data (cf. Wright and Masters 1982). Standard software for such analyses produces a person-specific estimated true score and standard error. The estimated true score becomes the outcome in equation (9), whereas the squared standard error becomes the level-1 variance in that equation. A limitation of this approach is that it does not easily accommodate item-level missing data.

A final approach is to use the item responses as the elemental data in the overall analysis. This might involve, for example, substituting equation (28) for equation (9) with the level-2 and level-3 models including the relevant covariates (as in equations (13) and (14)). Using this approach, the measurement error variances are estimated simultaneously with all other model parameters. We have, in fact, estimated this model. The results are essentially identical to those tabulated in this article. However, the computations are far more intense. Raudenbush and Sampson (1999) provide a detailed presentation of the integration of binary item response models into a hierarchical model.

Our general recommendation is to employ the most sensible available item response analysis to compute person-specific estimated true scores and standard errors. The level-1 variances of equation (9) are then assumed known and equal to the square of those standard errors. Such known level-1 variances are readily handled with available software for hierarchical models. For example, in using the HLM program (Bryk, Raudenbush, and Congdon 1996), the reciprocal of the squared standard error is defined as a weighting variable at level 1. The level-1 error variance is then constrained to unity. The resulting weighted analysis estimates the model specified by equations (9), (13), and (14).

NOTES

1. Conforming to common practice, we use the term *mediated effects* and the related terms *indirect effects* and *direct effects*. In many studies, a causal theory drives model specification and supports hypothesized direct and indirect effects, but nonexperimental research designs rarely support strong causal interpretations. Our use of the term *effects* is therefore not intended to imply a causal inference.

2. Commonly used packages include HLM (Bryk, Raudenbush, and Congdon 1996), MLN (Rasbash et al. 1995), PROC MIXED (SAS Institute 1996), and VARCL (Longford 1988).

3. In the applications we have encountered to date, the exogenous variables, X_k, can reasonably be assumed to be measured without error. For example, in our illustrative data, X_k are taken from the U.S. census, and the assumptions of reliable measurement seem reasonable. When X_k are assumed to be measured without error, it is convenient to work with the conditional distribution of Y_k and Z_k given X_k. It also simplifies the presentation to follow. However, information on measurement errors of X_k can readily be incorporated into the approach, in which case it is more convenient to work with the joint distribution of Y_k, X_k, and Z_k.

4. It is not clear that higher order approximations would be much of an improvement over these first-order approximations in light of the approximate nature of inference based on the information matrix in the context of ML. However, this topic is worthy of further investigation. For now, these techniques are best regarded as useful in reasonably large-sample contexts, cer-

tainly including the data in the example, based on 342 neighborhood clusters and over 7,000 survey respondents.

5. Within 80 of the 342 neighborhoods, approximately 6,000 young people have been sampled (independent of the community survey sample) and assessed for the first of eight annual waves of a longitudinal study of social development. Within these 80 neighborhoods, the community survey was designed to make 40 to 50 interviews per neighborhood available in order to maximize reliability of neighborhood measurement. Neighborhood measures collected as part of the community survey will produce explanatory variables for understanding the development of the longitudinal participants. The community survey was designed to produce approximately 20 interviews available in the remaining 262 nonsample neighborhoods so that, when a longitudinal sample member moves from a sample to a nonsample neighborhood within Chicago, significant useful information about that neighborhood will be available. As a result of this design, there is quite substantial variation in the reliability with which the 342 neighborhoods are assessed in the community survey. The analysis is designed to fully incorporate the information about this varying reliability.

6. Note that in this example, both Y_{jk} and Z_{jk} are assessed using data from the same reporter. Ideally, Y and Z would be measured on different reporters. This will be the case in the Project on Human Development in Chicago Neighborhoods' longitudinal study when Y will be measured on young people in the longitudinal study and Z will be taken from the independent sample of adults in the community survey. In other analyses, Sampson, Raudenbush, and Earls (1997) took Y from official crime statistics independent of the measurement of Z from the community survey. The statistical model proposed here can accommodate such analyses.

7. If the predictors in equation (2) are centered on their grand means, Y_k and Z_k will literally be adjusted means. In general, they are intercepts (expected values of Y_{jk} and Z_{jk} when the predictors are zero). We find that grand mean centering eases interpretation.

8. If predictors measured with normally distributed error are also viewed as having normally distributed random coefficients, the resulting marginal distribution will involve a product of two normals and will not be normal. Such a model presents special problems for estimation theory.

REFERENCES

Allison, Paul. 1995. "The Impact of Random Predictors on Comparisons of Coefficients Between Models: Comment on Clogg, Petkova, and Haritou." *American Journal of Sociology* 100:1294-1305.

Bollen, K. A. 1989. *Structural Equations With Latent Variables*. New York: Wiley.

Bowles, Samuel, and Herbert Gintis. 1996. "Productive Skills, Labor Discipline, and the Returns to Schooling." Department of Economics, University of Massachusetts, unpublished paper.

Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Bryk, Anthony S., Stephen W. Raudenbush, and Richard C. Congdon. 1996. *An Introduction to HLM: Computer Program and User's Guide, Version 2*. Chicago: Department of Education, University of Chicago.

Clogg, Clifford, Eva Petkova, and A. Haritou. 1995. "Symposium on Applied Regression: Statistical Methods for Comparing Regression Coefficients Between Models." *American Journal of Sociology* 100:1261-93.

Dempster, Arthur, Donald B. Rubin, and Robert C. Tsutakawa. 1981. "Estimation in Covariance Components Models." *Journal of the American Statistical Association* 76:341-53.

Goldstein, Harvey. 1995. *Multilevel Statistical Models*. 2d ed. New York: John Wiley & Sons.

Jennrich, Robert, and M. Schluchter. 1986. "Unbalanced Repeated-Measures Models With Structured Covariance Matrices." *Biometrics* 42:805-20.

Jöreskog, Karl, and Dag Sörbom. 1996. *LISREL 8: Structural Equation Modeling With the SIMPLILS Command Language*. Hove and London: Scientific Software International.

Lee, S. 1990. "Multilevel Analysis of Structural Equation Models." *Biometrika* 77:763-72.

Lee, Valerie, and Anthony S. Bryk. 1989. "A Multilevel Model of the Social Distribution of Educational Achievement." *Sociology of Education* 62:172-92.

Little, Roderick J., and Nathaniel Schenker. 1995. "Missing Data." Pp. 39-76 in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. New York: Plenum Press.

Longford, Nicholas T. 1987. "A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Models With Nested Random Effects." *Biometrika* 74:817-27.

———. 1988. *VARCL: Software for Variance Component Analysis of Data With Hierarchically Nested Random Effects*. Princeton, NJ: Educational Testing Service.

———. 1993. *Random Coefficient Models*. Oxford: Clarendon Press.

Mason, William, George Wong, and Barbara Entwisle. 1984. "Contextual Analysis Through the Multilevel Linear Model." Pp. 72-103 in *Sociological Methodology 1983-1984*, edited by S. Leinhardt. San Francisco: Jossey-Bass.

McDonald, Roderick. 1993. "A General Model for Two-Level Data With Responses Missing at Random." *Psychometrika* 4:575-85.

———. 1994. "The Bilevel Reticular Action Model for Path Analysis and Latent Variables." *Sociological Methodology and Research* 22:399-413.

McDonald, Roderick, and Harvey Goldstein. 1989. "Balanced Versus Unbalanced Designs for Linear Structural Relations in Two-Level Data." *British Journal of Mathematical and Statistical Psychology* 42:215-32.

Muthen, Bengt. 1990. "Multilevel Structural Equation Modeling." Paper presented at the annual meeting of the American Educational Research Association, April, Boston, MA.

Powell, Kenneth E., Linda Dahlberg, Jennifer Friday, James Mercy, Timothy Thornton, and Shaunette Crawford. 1996. "Prevention of Youth Violence: Rationale and Characteristics of 15 Evaluation Projects." *American Journal of Preventive Medicine* 12 (5): 3-12.

Rasbash, Jonathan, Ming Yang, Geoffrey Woodhouse, and Harvey Goldstein. 1995. *MLN: Command Reference Guide*. London: Institute of Education.

Raudenbush, Stephen W. 1995. "Hierarchical Linear Models to Study the Effects of Social Context on Development." Pp. 165-201 in *The Analysis of Change*, edited by J. Gottman. Hillsdale, NJ: Lawrence Erlbaum.

Raudenbush, Stephen W., Brian Rowan, and Sang Jin Kang. 1991. "A Multilevel, Multivariate Model for Studying School Climate in Secondary Schools With Estimation via the EM Algorithm." *Journal of Educational Statistics* 16:295-330.

Raudenbush, Stephen W., and Robert Sampson. 1999. "Ecometrics: Toward a Science of Assessing Ecological Settings, With Application to the Systematic Social Observation of Neighborhoods." *Sociological Methodology* 29 (1): 1-41.

Rivera-Batiz, F. 1992. "Quantitative Literacy and the Likelihood of Employment Among Young Adults in the United States." *Journal of Human Resources* 27:313-28.

Sampson, Robert, and W. Groves. 1989. "Community Structure and Crime: Testing Social-Disorganization Theory." *American Journal of Sociology* 94:774-802.

Sampson, Robert, and Janet Lauritsen. 1994. “Violent Victimization and Offending: Individual, Situational and Community-Level Risk Factors.” Pp. 1-114 in *Understanding Preventing Violence Social Influences*, edited by Albert Reiss and J. Roth. Washington, DC: National Academy Press.

Sampson, Robert, Stephen Raudenbush, and Felton Earls. 1997. “Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy.” *Science* 277:918-24.

SAS Institute. 1996. *PROC MIXED: Technical Report P-229*. Cary, NC: SAS/STAT.

Schafer, Joseph. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Woodhouse, Geoffrey, Ming Yang, Harvey Goldstein, and Jonathan Rasbash. 1996. “Adjusting for Measurement Error in Multilevel Analysis.” *Journal of the Royal Statistical Society A* 159:201-12.

Wright, Benjamin, and G. Masters. 1982. *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.

Yuan, Ke-Hai, and Peter Bentler. 1997. “Mean and Covariance Structure Analysis: Theoretical and Practical Improvements.” *Journal of the American Statistical Association* 92:767-74.

Stephen W. Raudenbush is a professor in the School of Education and a senior research scientist in the Survey Research Center at the University of Michigan. His major research interests include the design and analysis of multilevel and longitudinal research and the assessment of social settings in which young people develop.

Robert Sampson is the Lucy Flower Professor of Sociology at the University of Chicago and a research fellow at the American Bar Foundation. His research interests include criminology, community/urban sociology, and the life course.