

Multilevel Modeling of Individual and Group Level Mediated Effects

Jennifer L. Krull
University of Missouri-Columbia

David P. MacKinnon
Arizona State University

This article combines procedures for single-level mediational analysis with multilevel modeling techniques in order to appropriately test mediational effects in clustered data. A simulation study compared the performance of these multilevel mediational models with that of single-level mediational models in clustered data with individual- or group-level initial independent variables, individual- or group-level mediators, and individual level outcomes. The standard errors of mediated effects from the multilevel solution were generally accurate, while those from the single-level procedure were downwardly biased, often by 20% or more. The multilevel advantage was greatest in those situations involving group-level variables, larger group sizes, and higher intraclass correlations in mediator and outcome variables. Multilevel mediational modeling methods were also applied to data from a preventive intervention designed to reduce intentions to use steroids among players on high school football teams. This example illustrates differences between single-level and multilevel mediational modeling in real-world clustered data and shows how the multilevel technique may lead to more accurate results.

Mediational analysis is a method that can help researchers understand the mechanisms underlying the phenomena they study. The basic mediational framework involves a three variable system in which an initial independent variable affects a mediational variable, which, in turn, affects an outcome variable (Baron & Kenny, 1986). The aim of mediational analysis is to determine whether the relation between the initial variable and the outcome is due, wholly or in part, to the mediator. Mediational analysis is applicable across a wide range of experimental and non-experimental

This research was supported by grant T32 MH18387 from the National Institute of Mental Health and grant 5-R01-DA09757-02 from the National Institute on Drug Abuse. Collection of the ATLAS data was supported by National Institute on Drug Abuse grant 5-R01-DA07356-05.

Correspondence concerning this article should be addressed to Jennifer L. Krull, Department of Psychological Sciences, 210 McAlester Hall, University of Missouri, Columbia, MO 65211.

research situations, including both process-oriented laboratory research and field studies.

Mediational analysis is particularly useful in prevention research, where it allows researchers to explore the mechanisms through which experimental interventions have their effects. Often preventive interventions are designed to affect certain key process variables (i.e., mediators) which are believed to affect the outcome measure of interest. For example, drug use intervention programs often target mediational variables such as knowledge of the effects of particular drugs, beliefs about the prevalence of drug use, and resistance skills (Hansen, 1992; MacKinnon & Dwyer, 1993). Changes in these mediators are believed to affect change in intentions to use drugs as well as change in actual drug use.

Often preventive interventions are delivered within an organizational system, such as a school system, which provides ready access to intact groups of subjects. With students nested within classrooms and schools, the data are inherently clustered. Clustered data is frequently encountered in other areas of psychological research as well. For example, health researchers may study the outcomes of patients nested within clinics or hospitals, industrial/organizational researchers may be interested in individual workers nested within work groups or companies, and survey researchers may encounter respondents nested within various geographical areas, such as census tracts, neighborhoods, or countries. It is likely that the students who attend a particular school or the individuals who reside within a particular geographic area are more similar to each other than a randomly selected group of students or individuals would be. Such similarity may be due to shared group experiences, reciprocal influence resulting from group interaction, or non-randomly distributed background variables. When modeling data gathered from these clustered individuals, the within-group homogeneity (indexed by the intraclass correlation or ICC) results in positively correlated error terms among the individuals within a particular group. In such a situation (i.e., with positive ICC), mediational analysis suffers from the same difficulty as other traditional analytic techniques, namely the unit of analysis problem (Palmer, Graham, White, & Hansen, 1998).

Because traditional analyses are not designed to accommodate clustered data, their application requires the researcher to make a decision about whether the analysis should be based on individual level responses, ignoring the clustering, or on aggregate measures for each group, eliminating the individual scores. In other words, should the unit of the analysis be the individual or the group? If the individual was chosen as the unit of analysis, the analysis would proceed as if the data were unclustered and any group

level variables were, in fact, individual level variables. However, with positive ICC frequently found in clustered data, the correlated errors among the individuals within a group violate the independent observations assumption of ordinary least squares (OLS) estimation, resulting in downwardly biased standard error estimates, overly large test statistics, and inflated Type I error rates (Barcikowski, 1981; Moulton, 1986; Scariano & Davenport, 1987; Scott & Holt, 1982; Walsh, 1947). If the analysis were to proceed at the group level, all individual level data would be aggregated into group means, and groups, not individuals, would be treated as the observations in the analysis (see, for an example, MacKinnon et al., 1991). Though this eliminates the correlation among individual error terms, this approach is not ideal for a number of reasons. First, the power of statistical tests is reduced because the number of observations is limited to the number of groups, rather than the number of individuals, and the degrees of freedom for the analysis are correspondingly decreased. Second, this procedure discards the individual variability in scores, which may comprise the majority of the variability present in the data (de Leeuw, 1992). In addition, primary interest in prevention studies typically centers on the individual, but the group level relationships explored by an aggregate analysis do not necessarily parallel individual level effects, making inferences about individual behavior potentially misleading (Robinson, 1950; Pedhazur, 1982, p. 529). Thus conducting an analysis of clustered data by aggregating all individual level information to the group level is often not an optimal solution to the correlated error problem.

Multilevel modeling was developed in response to the challenge of appropriately analyzing clustered data. This technique preserves the original data structure (i.e., individual level variables need not be aggregated to group means) while explicitly modeling the within-group homogeneity of errors by allowing the estimation of error terms for both the individual and the group. Because of the complex structure of the model and the nature of the error terms, multilevel models are estimated using iterative Empirical Bayes/maximum likelihood (EB/ML) techniques, rather than the OLS methods typically employed to estimate the parameters of single-level models. The standard error estimates for a multilevel model are more accurate than those for a single-level individual-as-unit-of-analysis model. In addition to the correction of standard error estimates and the more appropriate significance tests that result, multilevel models also provide other advantages over traditional analytic techniques. Prominent among these is the ability to simultaneously examine the effects of variables at both individual and group levels, as well as possible cross-level interaction effects (Bryk & Raudenbush, 1992).

In the field of prevention research, school-based interventions are often randomized at the classroom or school level (i.e., intact classes or schools are assigned to either control or intervention conditions), yet the outcome measures of primary interest involve characteristics or behaviors of individual students (e.g., drug use). Prevention researchers have recognized the appropriateness of multilevel models for estimating the overall effect of the group level intervention on individual level drug use outcomes (e.g., Hedeker, Gibbons, & Flay, 1994; Murray, 1998; Palmer et al., 1998). Multilevel modeling corrects the standard error associated with the estimate of the program effect, resulting in a more accurate test of its significance. In addition to this primarily statistical advantage, prevention researchers have also recognized that the multilevel framework provides an opportunity to examine possible interactions between the intervention and both individual and group level variables. For example, Kreft (1997) used multilevel models to examine individual rebelliousness as a possible moderator of the effect of a class-level drug use intervention on individual alcohol and cigarette use. Other researchers speculate that the effectiveness of a classroom-based alcohol and drug use prevention program may depend on the receptiveness of the individual student, the ability of the teacher to implement the program in the classroom, and community resources (Palmer et al., 1998). The multilevel framework allows researchers to consider such moderational effects at any level of their data, knowing that appropriate statistical tests are available.

In addition to multilevel moderation, the ability of the multilevel framework to examine the effects of both individual and group level variables also opens up the possibility of conducting analyses to examine multilevel mediation. Just as moderational effects can involve interactions between variables at either level, the processes involved in mediation may also operate at either (or both) of the levels of a clustered dataset. For example, examination of the processes which underlie a drug use prevention program may involve testing the mediational role of both individual variables, such as individual knowledge of the effects of various drugs, and group level variables, such the group norm regarding drug use. Multilevel modeling techniques can be incorporated into the framework of mediational analysis to provide tests of the mediated effects in models of this general type, with data nested within clusters and/or key variables measured at different levels.

The basic prerequisites for the appropriate application of multilevel mediational analysis include (a) clustered data with positive ICC, and (b) a proposed three-variable mediational model in which the outcome variable is measured at the lowest level of the data (in the examples presented here, the individual level). Additionally, the structure of multilevel models dictates that each link in the mediational chain may involve a variable affecting another

measured at the same level or at a lower level, but not at a higher level. In other words, a group level variable may affect either a group level variable or an individual level variable; an individual level variable may affect another individual level variable, but not a group level variable. The simplest situation that meets these criteria would involve a dataset in which individuals are nested within intact groups and a hypothesis involving a proposed mediational chain in which all three variables (initial, mediator, and outcome) are individual characteristics or behaviors. Such a case is pictured in the top panel of Figure 1. The model is labeled $1 \rightarrow 1 \rightarrow 1$, the three numbers indicating that the initial (X_{ij}), mediator (M_{ij}), and outcome (Y_{ij}) variables in the model, respectively, are measured at Level 1, or the lowest level of the data. The ij subscript on each variable indicates that the variable can take on a unique value for each individual i within each group j . In this simplest case with all three variables measured at the same level, it is only the clustered nature of the data that requires multilevel modeling to appropriately model the error structure. If the initial variable in the chain were continuous, this model might apply to process-oriented survey research in which the predictor variable was a construct measured on a Likert-type scale or to an experimental study where a treatment was administered at a number of different dosing levels. If the initial variable were dichotomous, this model might apply to an experimental study in which individual subjects are assigned to control or treatment conditions or to an observational study in which the measured initial variable can take on only two different values.

Mediational models that include group level initial or mediator variables may also be analyzed with multilevel mediational models. Consider, for example, a $2 \rightarrow 1 \rightarrow 1$ model, as depicted in the center panel of Figure 1. Here, an initial level 2 variable (X_j), representing a characteristic of the group, affects an individual level mediator (M_{ij}), which, in turn, affects an individual level outcome (Y_{ij}). The single subscript j on the X variable indicates that this variable may take on a unique value for each group j . Such a scenario fits the typical preventive evaluation example with a group-based intervention designed to change individual knowledge or perceptions in order to reduce individual drug use. A combination of single-level and multilevel modeling also makes possible the analysis of a $2 \rightarrow 2 \rightarrow 1$ model, in which both the initial (X_j) and mediator (M_j) variables are measured at the group level, but the outcome (Y_{ij}) is an individual behavior. Such a model is depicted in the bottom panel of Figure 1. An example of such a model is provided by Sampson, Raudenbush, and Earls (1997), who hypothesized that neighborhood level measures of social composition (which reflect economic disadvantage and immigrant concentration) affect neighborhood level measures of collective efficacy (social cohesion and informal social control), which, in turn, influence individual

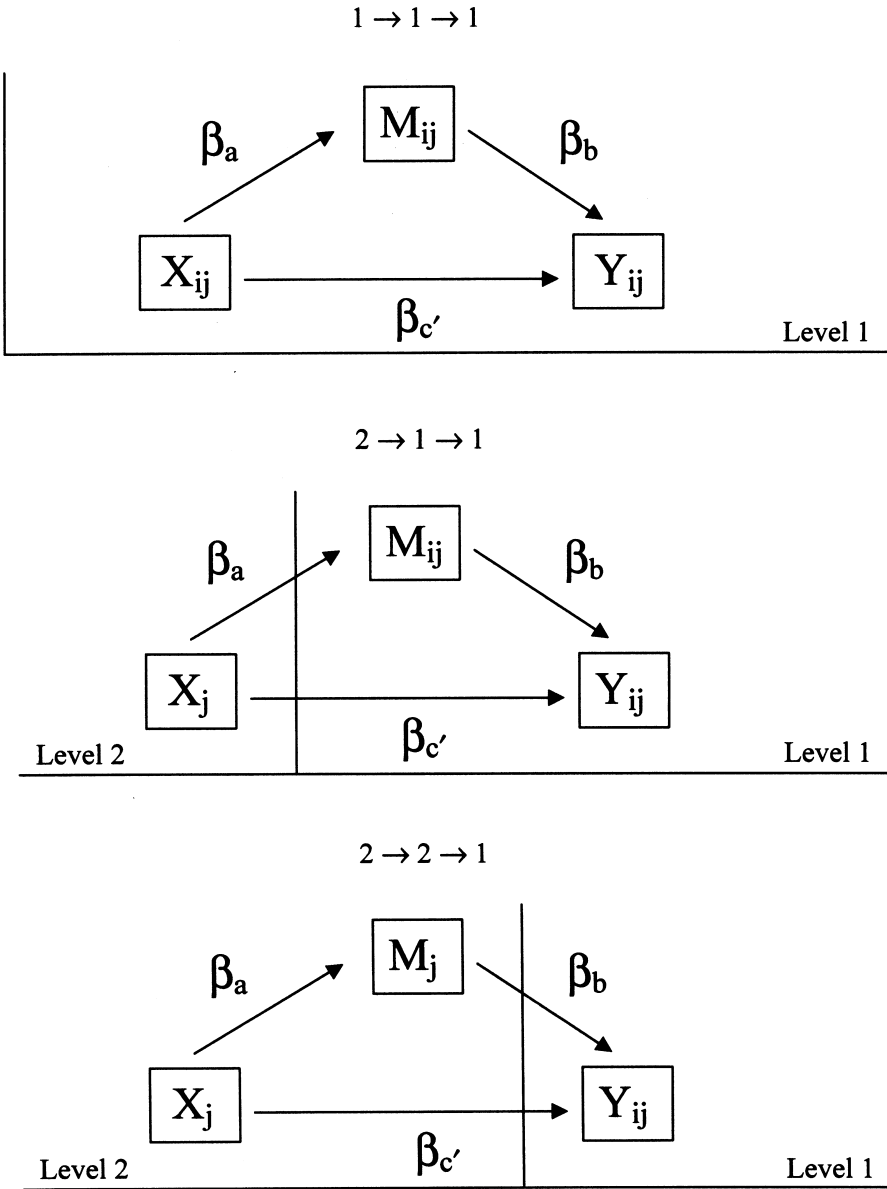


Figure 1
Three Types of Mediation Models

level measures of violence (such as perceived level of neighborhood violence and individual violent victimization).

The analysis of multilevel mediational effects is somewhat complicated by the fact that many variables can be conceptualized at more than one level. Any individual level measure can be aggregated to the group level, simply by taking the mean for each group. Effects involving the variable may operate at either or both levels. Moreover, individual and aggregate measures of the same variable may reflect quite different constructs at the various levels. For example, individual student level responses regarding parental occupation and education may serve as indicators of home background and reflect parental commitment to the student's learning, while school level aggregates of the same responses more likely indicate the wealth and urbanism of the community, which may determine the level of school resources (Burstein, 1980; 1985). In general, individual level variables tend to be more psychological in nature than group aggregates, which may be more indicative of organizational or normative aspects of the environment. Aggregate measures may also represent contextual influences, which can operate differently than the individual measures on which they were based. For example, rebelliousness as an individual characteristic may encourage risk-taking and make an intervention less likely to be effective. Moreover, a group with a high average level of rebelliousness may create an environment in which discipline issues would make program delivery difficult, decreasing program effectiveness even for non-rebellious individuals (Palmer et al., 1998).

Multilevel mediational modeling is a flexible technique which allows researchers to appropriately test mediated effects in clustered datasets and can accommodate initial and mediator variables measured at either the group or the individual level. The basic model can be extended to include multiple mediators operating at different levels, which could potentially disentangle the mediational effects of individual and aggregate mediational effects of the same variable. The following section of this article describes in detail the general structure of multilevel mediational models as a combination of single-level mediational analysis procedures and multilevel modeling techniques. The performance of multilevel mediational models relative to single-level mediational models is then evaluated in a simulation study. Finally, multilevel mediational modeling techniques are applied to data from a steroid use prevention study in order to illustrate the differences between single-level and multilevel mediational approaches in real-world data.

Mediational Analysis Procedures

In single-level mediational models, algebraically equivalent point estimates of the mediated effect can be calculated in two ways (MacKinnon, Warsi, & Dwyer, 1995), each of which involves the estimation of two regression equations. The straightforward reformulation of these regression equations in the multilevel framework allows for the analogous estimation of mediated effects in clustered multilevel data.

The first method for obtaining a point estimate of the mediated effect for single-level data requires the estimation of a regression equation predicting the outcome measure (Y_i) from the initial variable (X_i)

$$(1) \quad Y_i = \beta_0 + \beta_c X_i + r_i$$

and the estimation of a regression equation predicting the outcome measure from the initial variable and the mediator (M_i)

$$(2) \quad Y_i = \beta_0 + \beta_c X_i + \beta_b M_i + r_i$$

The difference between the estimates of the coefficients associated with the initial variable (X_i) in these two equations, $\hat{\beta}_c - \hat{\beta}_{c'}$, estimates the mediated effect as the extent to which the mediator accounts for the relationship between the initial and outcome variables (Judd & Kenny, 1981).

The second method for calculating point estimates of mediated effects requires estimating a regression equation predicting the mediator from the initial variable

$$(3) \quad M_i = \beta_0 + \beta_a X_i + r_i$$

and a regression equation identical to Equation 2 above predicting the outcome measure from the initial variable and the mediator

$$(4) \quad Y_i = \beta_0 + \beta_{c'} X_i + \beta_b M_i + r_i$$

The product $\hat{\beta}_a \hat{\beta}_b$ is a second point estimate of the mediated effect, which evaluates the extent to which the initial variable affects the mediator and the extent to which the mediator, in turn, affects the outcome. In single-level models, the two expressions, $\hat{\beta}_c - \hat{\beta}_{c'}$ and $\hat{\beta}_a \hat{\beta}_b$, produce algebraically equivalent estimates of the mediated effect (MacKinnon, Warsi, & Dwyer, 1995).

In single-level mediational models, methods for estimating the variance of a product of random variables are often used to estimate the standard error of the mediated effect for use in interval estimation and significance testing (Sobel, 1982). A number of such methods exist (e.g., Goodman, 1960; Mood, Graybill, & Boes, 1974), and previous work involving both single-level and multilevel mediational models suggests that the expression for the square root of the asymptotic variance based on the multivariate delta method (Sobel, 1982; 1986) or the first order Taylor series

$$(5) \quad \hat{s}_{\beta_a\beta_b} = \sqrt{\hat{s}_{\beta_a}^2 \hat{\beta}_b^2 + \hat{s}_{\beta_b}^2 \hat{\beta}_a^2}$$

performs well across a range of sample sizes (Krull & MacKinnon, 1999; MacKinnon, Warsi, & Dwyer, 1995). The expression within the radical may be derived by pre- and post-multiplying the covariance matrix of the random variables (β_a and β_b) by the partial derivatives of the desired function ($\beta_a \beta_b$) of the random variables, or by using Taylor series expansion to develop a polynomial approximation of the desired function and then calculating the variance of this polynomial. The estimates used in calculating this result may be obtained from regression analyses of Equations 2 and 3, simply adding the squared estimates of the effects β_a and β_b , each multiplied by the squared standard error of the other.

Each of the three unique single-level mediational equations (Equations 1 through 3) may be recast as a multilevel equation, generating multilevel coefficients and standard error estimates that may be used to calculate multilevel $\hat{\beta}_c - \hat{\beta}_c'$ or $\hat{\beta}_a \hat{\beta}_b$ estimates of the mediated effect, as well as a multilevel approximation for the standard error of the mediated effect, using the formula provided in Equation 5. Table 1 shows the multilevel reexpressions of each of the three unique single level equations for three different types of models. The simplest case is the 1 → 1 → 1 model in which all variables are measured at the individual level. Such a model would be applicable, for example, to a experimental intervention study where students within classes (rather than the classes) were randomly assigned to treatments believed to change individual perceptions about a particular behavior (e.g., smoking) which would then affect the extent to which the students engaged in the behavior. The single-level and multilevel mediational equations for this type of model are shown in the top section of Table 1. The multilevel specification for Equation 1, and for Equations 2 and 3 as well, includes an individual level (Level 1) equation and a related group level (Level 2) equation. The individual level equation is very similar to the single-

Table 1

Single-level and Multilevel Equations for Mediation Analysis for the Three Types of Models

		<u>Analysis Type</u>	
		<u>Single-Level</u>	<u>Multilevel</u>
$1 \rightarrow 1 \rightarrow 1$			
Equation 1	$Y_{ij} = \beta_0 + \beta_c X_{ij} + r_{ij}$		Level 1: $Y_{ij} = \beta_{0j} + \beta_c X_{ij} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$
Equation 2	$Y_{ij} = \beta_0 + \beta_c X_{ij} + \beta_b M_{ij} + r_{ij}$		Level 1: $Y_{ij} = \beta_{0j} + \beta_c X_{ij} + \beta_b M_{ij} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$
Equation 3	$M_{ij} = \beta_0 + \beta_a X_{ij} + r_{ij}$		Level 1: $M_{ij} = \beta_{0j} + \beta_a X_{ij} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$
$2 \rightarrow 1 \rightarrow 1$			
Equation 1	$Y_{ij} = \beta_0 + \beta_c X_j + r_{ij}$		Level 1: $Y_{ij} = \beta_{0j} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + \gamma_c X_j + u_{0j}$
Equation 2	$Y_{ij} = \beta_0 + \beta_c X_j + \beta_b M_{ij} + r_{ij}$		Level 1: $Y_{ij} = \beta_{0j} + \beta_b M_{ij} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + \gamma_c X_j + u_{0j}$
Equation 3	$M_{ij} = \beta_0 + \beta_a X_j + r_{ij}$		Level 1: $M_{ij} = \beta_{0j} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + \gamma_a X_j + u_{0j}$
$2 \rightarrow 2 \rightarrow 1$			
Equation 1	$Y_{ij} = \beta_0 + \beta_c X_j + r_{ij}$		Level 1: $Y_{ij} = \beta_{0j} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + \gamma_c X_j + u_{0j}$
Equation 2	$Y_{ij} = \beta_0 + \beta_c X_j + \beta_b M_j + r_{ij}$		Level 1: $Y_{ij} = \beta_{0j} + r_{ij}$ Level 2: $\beta_{0j} = \gamma_{00} + \gamma_c X_j + \gamma_b M_j + u_{0j}$
Equation 3	$M_j = \beta_0 + \beta_a X_j + r_j$		

level equation, with the individual level outcome Y_{ij} predicted by an intercept term, the individual level initial variable X_{ij} , and an individual level error term. However, the intercept in the multilevel specification is indexed by the subscript j , which indicates that the intercept is treated as a random coefficient and each group j may have a unique intercept value. Each group intercept is then predicted in the group level equation from an overall intercept term, γ_{00} , and a group level error term, u_{0j} . It is this second error term that allows the multilevel model to accommodate the correlated error

structure inherent in clustered data. Equations 2 and 3 are restructured similarly. In each case, the intercept is specified as a random coefficient, and error terms are included at each level of the model.

The 2 → 1 → 1 model differs in that the initial variable in the mediational chain is a group level variable. Such a data structure is frequently encountered in prevention research in which intact groups are randomly assigned to treatment conditions designed to change individual mediators (e.g., perceptions, knowledge, beliefs about drug use) which in turn affect individual outcomes (e.g., drug use behaviors). The single-level and multilevel mediational equations for this model are shown in the center section of Table 1. In the multilevel specification, the group level initial variable is incorporated into the group level equations. For example, in Equation 1, the X_j variable is included as a predictor of the group intercept in the Level 2 equation. The associated coefficient is $\hat{\gamma}_c$, rather than the $\hat{\beta}_c$ of the single-level specification, but this notational change does not impact the way in which the coefficient is used in the mediational analysis. Inclusion of the X_j predictor in the group level specification of Equation 2 produces the multilevel estimate of $\hat{\gamma}_c$, and $\hat{\gamma}_c - \hat{\gamma}_c$ is the multilevel equivalent of the $\hat{\beta}_c - \hat{\beta}_c$ mediated effect. Similarly, Equation 3 produces the multilevel coefficient $\hat{\gamma}_a$, which can be substituted for $\hat{\beta}_a$ in calculating the multilevel mediated effect $\hat{\gamma}_a \hat{\beta}_b$, analogous to the single-level $\hat{\beta}_a \hat{\beta}_b$ estimate.

The 2 → 2 → 1 model includes both initial and mediator variables at the group level. This might represent intact groups randomized to treatments designed to affect global measures of group functioning, which would then impact individual behavior. For such a model, Equation 1 is formulated as for the previous model, with the X_j variable predicting the group intercept in the Level 2 equation. Equation 2 includes both the group level initial variable X_j and the group level mediator M_j in the group level equation. Equation 3, which predicts the mediator from the initial variable, involves only group level variables. Therefore, a single-level (i.e., group level only) specification is appropriate, and estimates of $\hat{\beta}_a$ obtained via OLS estimation should be used in calculating the multilevel $\hat{\beta}_a \hat{\gamma}_b$ estimate of the mediated effect.

In this set of multilevel equations, only the intercept term has been specified as a random coefficient. It would be possible in those models involving an individual level mediator to specify the β coefficient associated with the mediator as random as well and predict this variation in the relationship between mediator and outcome in an additional group level equation. However, the simulation and example analyses reported here are based on the specifications in Table 1 with the single random intercept coefficient, in order to provide the most direct comparison between fixed coefficient single-level and multilevel mediational models. Moreover, the

equations are formulated using uncentered variables. In multilevel models, centering decisions impact the interpretation of intercepts, the variance of random intercepts across groups, and the covariance of intercepts with random slopes (Bryk & Raudenbush, 1992, p. 28). In these random-intercept-only models, the interpretation of these particular parameters is not crucial to obtaining and interpreting estimates of β or γ coefficients to be used in calculating estimates of mediated effects, and thus it is possible to work with the variables in their original form.

The $\hat{\beta}_c - \hat{\beta}_{c'}$ and $\hat{\beta}_a\hat{\beta}_b$ estimates of the mediated effect, algebraically equivalent in single-level models, are not exactly equivalent in multilevel models (Krull & MacKinnon, 1999). This is due to the fact that the weighting matrix used in estimating the fixed effects in the multilevel equations is not necessarily identical for each of the three equations. The non-equivalence, however, is unlikely to be problematic because the discrepancy between the two estimates is typically small and unsystematic, essentially vanishing at larger sample sizes (Krull & MacKinnon, 1999).¹ Decisions about which estimate to use may, therefore, be based on other aspects of estimator performance. For example, in a model including more than one mediated effect, use of the $\hat{\beta}_a\hat{\beta}_b$ estimator provides a separate estimate of each of the mediated effects, while the $\hat{\beta}_c - \hat{\beta}_{c'}$ estimator gives only an estimate of the total mediated effect. Therefore, the $\hat{\beta}_a\hat{\beta}_b$ estimate of the mediated effect is used in both the simulation study and the real-world prevention example presented below. Moreover, the $\hat{\beta}_a\hat{\beta}_b$ estimate of the mediated effect lends itself readily to subsequent examination of its component $\hat{\beta}_a$ and $\hat{\beta}_b$ paths. Such analyses can be particularly informative when statistical tests show that the mediated effect is not significantly different from zero. For example, in the evaluation of a prevention program, a non-significant mediated effect might be traced back to a non-significant $\hat{\beta}_a$ path linking the initial (program) and mediator variables, indicating an “action theory failure” in which the program has failed to affect the hypothesized process variable (Chen, 1990). Alternately, a non-significant mediated effect might be traced back to a non-significant $\hat{\beta}_b$ path, indicating a “conceptual theory failure” in which the process variable, though changed by the program, has had no impact on the outcome. In this case, the theory which links the process variable to the outcome would be suspect. Corrective action would differ for the two cases:

¹ It should be noted, however, that this equivalence may not necessarily hold for non-linear models (e.g., survival analyses, logistic regression). In some cases, rescaling of variances may be sufficient to restore the equivalence (see MacKinnon & Dwyer, 1993). In other cases where the difference between the two mediated effect estimates is not so easily resolved, the analyst may need to consider the logic underlying the two expressions and decide which more closely parallels the research question in order to make the decision about which mediated effect estimator to use.

program restructuring may be necessary to deal with action theory failure, while conceptual theory failure would suggest a search for better replacement theory.

Simulation Study

A simulation study was conducted to compare the performance of multilevel and single-level mediational analyses in estimating and testing mediated effects in clustered data and to determine how such performance might be related to other factors, including characteristics of the clustered datasets themselves and the true parameters defining the mediational relationship. The SAS programming language was used to generate simulated data sets to represent mediational chains in which an initial variable X affected a mediator M , which, in turn, affected an outcome variable Y . Data were generated to have multilevel structure (i.e., with observations drawn from individual units nested within distinct groups), with the outcome variable structured to represent a variable measured for each individual subject within each group. Models were simulated to reflect $1 \rightarrow 1 \rightarrow 1$, $2 \rightarrow 1 \rightarrow 1$, and $2 \rightarrow 2 \rightarrow 1$ data structures. For each data structure, separate sets of models were simulated with dichotomous and continuous initial variables beginning the mediational chain.

Six factors were systematically varied in the simulations. Two aspects of sample size, the number of groups and group size, were manipulated such that the simulated data sets were comprised of 10, 20, 30, 50, 100, or 200 groups of either small or moderate size. For the small group size conditions, half of the groups were comprised of 5 individuals, the other half of 10 individuals. For the moderate group size conditions, half of the groups had 20 individuals, the other half had 30 individuals in each group.

The proportion of residual variance in the simulated mediator and outcome variables due to between-group variability was also varied. In $1 \rightarrow 1 \rightarrow 1$ and $2 \rightarrow 1 \rightarrow 1$ models, the residual ICCs for the mediator (ICC_M) and outcome (ICC_Y) variables were adjusted by changing the variability of the group error terms, u_{0j} , relative to the variability of the individual error terms, r_{ij} , which was set equal to 1.0. Values were chosen so that the proportion of residual variance due to between-group variability (ICC), calculated as $\text{Var}(u_{0j}) / [\text{Var}(r_{ij}) + \text{Var}(u_{0j})]$, was equal to .05 or .15. These values were chosen to be representative of actual proportions that might be observed in intervention studies in a number of areas, including psychology, education, and family studies. In $2 \rightarrow 2 \rightarrow 1$ models, manipulating the variability of the group error term for the M variable simply changed the total variance for the mediator rather than changing "ICC", since there was no

individual level variability in this group level variable. ICC of the initial X_{ij} variable was simulated to be a relatively high .30 in the continuous $1 \rightarrow 1 \rightarrow 1$ model, in order to produce null model ICCs (calculated without any predictors in the model) roughly comparable to those of the other models. The initial X_{ij} variable was simulated with zero ICC in the dichotomous $1 \rightarrow 1 \rightarrow 1$ model, consistent with the treatment of this variable as random assignment to an experimental condition.

Finally, the true values of the β_a and β_b parameters which comprise the mediated effect were also systematically varied. In the simulated model, the β_a parameter (representing the effect of the intervention on the mediator) could take on values of .3 or .6, and the β_b parameter (representing the effect of the mediator on the outcome) could take on values of .1 or .4. The β_c parameter (representing the direct effect of the intervention on the outcome) was set to a constant value of .1. These parameter values of β_a and β_b are similar in magnitude to those used in previous simulation studies of mediational models (e.g., Krull & MacKinnon, 1999; MacKinnon, Warsi, & Dwyer, 1995), and the true values of the mediated effect $\beta_a\beta_b$ (which range from .03 to .24) reflect a range of values which might be encountered in preventive intervention studies.

All possible combinations of the 6 values for the number of groups in the analysis, the 2 group sizes, the 2 true values of β_a , the 2 true values of β_b , the 2 ICC_M values, and the 2 ICC_Y values were examined. Five hundred replications were conducted for each of the 192 conditions, producing 96,000 simulated data sets for each of the six model types.

For each data set, results from the OLS estimation of single-level Equations 2 and 3 (using SAS PROC REG) and the corresponding EB/ML estimation of multilevel equations from Table 1 (using SAS PROC MIXED) were used to form the single-level and multilevel $\hat{\beta}_a\hat{\beta}_b$ estimates of the mediated effect. These results were also used to calculate multivariate delta method/first order Taylor series approximations (Equation 5) for the standard errors of both single-level and multilevel mediated effects.

Relative bias in mediated effect estimates and standard error approximations were examined to determine the conditions under which the advantages of multilevel mediational analysis would be most evident. Relative bias was calculated using the equation

$$(6) \quad RB = \frac{\hat{w} - w}{w},$$

where \hat{w} is an estimate of the mediated effect or its standard error, and w is the true parameter value or an approximation of the true value when the precise value is not known. True values of the mediated effect were the values $\beta_a\beta_b = .03, .06, .12$ or $.24$ used to generate the simulated data. Empirical standard errors, calculated as the standard deviation of mediated effect estimates across the 500 replications in a given condition, were used as approximations to the true values in calculating the relative bias of the estimated standard errors.

Following the recommendation of Hauck and Anderson (1984) that simulation studies be analyzed using the same analytical tools as any other experimental study, factorial analyses of variance were conducted on the relative bias values for mediated effect estimates and standard errors, using number of groups, group size, ICC_M , ICC_Y , and true values of β_a and β_b parameters as predictors. Analysis type (single-level vs. multilevel) was treated as an additional repeated measures factor in these analyses, which were carried out using a multivariate analysis of variance approach to handle the repeated measures data. In order to identify the conditions under which multilevel estimation of the mediated effect is likely to show the greatest advantage over single-level estimation, interactions between the repeated analysis type factor and the six manipulated factors were examined. Due to the large number of observations, standard probability values were not particularly helpful in identifying meaningful effects. Therefore, the multivariate correlation ratio, a measure of multivariate effect size, was calculated for each effect using the equation

$$(7) \quad \eta_{mult}^2 = 1 - \Lambda$$

where Λ is the Wilks' likelihood-ratio criterion for the effect in question (Tatsuoka, 1988, p. 97, 288). An η_{mult}^2 value greater than or equal to .05 was used to identify the most important interactions in this study.

Simulation Results

The multiple factors and large number of conditions in the simulation study make it difficult to adequately represent the results concisely in tabular form. The following sections of text describe the results of the analyses of simulated data. Complete tables of relative bias values are available by request from the first author. Analyses of the relative bias in single-level (OLS) and multilevel (EB/ML) estimates of the mediated effect showed essentially no bias in either estimator, regardless of the data structure or the continuous or dichotomous nature of the initial variable. Analyses of the

relative bias in the standard error estimate of the single-level and multilevel mediated effects are discussed below. Results involving sample size and ICC are presented separately for those models involving dichotomous and continuous initial variables. Where significant differences between the two analysis types are found, an attempt is made to determine whether these results are due to effects involving β_a or β_b by examining the amount of relative bias associated with each path. This is followed by results regarding the true value of the β_a and β_b parameters and a summary of general patterns of results found across the six model types.

Effects of Sample Size and ICC

Dichotomous Initial Variable Models

Three models were examined in which a dichotomous initial variable served as the initial variable in the mediational chain.

1 → 1 → 1. In the model in which the mediational chain involved three individual level variables, single-level modeling resulted in slight underestimation the standard error of the mediated effect. The extent of this underestimation was greatest for conditions involving higher levels of ICC_M [$\Lambda = .9386, F_{(1,95808)} = 6269.75, p < .0001, \eta^2 = .0614$], higher levels of ICC_Y [$\Lambda = .9366, F_{(1,95808)} = 6483.41, p < .0001, \eta^2 = .0634$], and larger group sizes [$\Lambda = .8863, F_{(1,95808)} = 12286.11, p < .0001, \eta^2 = .1137$]. Even under the most extreme combination of conditions for this model, the relative bias in the single-level estimates of the standard error of the mediated effect was only about 10% of the true value of the standard error. For those conditions that showed such bias, it could be traced back to underestimation of the standard error of the β_b effect. Estimates of the standard error of β_a were essentially unbiased, as were all estimates obtained using multilevel techniques.

2 → 1 → 1. For the model in which a dichotomous group level predictor affected an individual level mediator, ICC_M [$\Lambda = .7103, F_{(1,95808)} = 39084.29, p < .0001, \eta^2 = .2897$] and group size [$\Lambda = .5713, F_{(1,95808)} = 71904.36, p < .0001, \eta^2 = .4287$] were again factors which affected the extent to which multilevel modeling showed an advantage over single-level modeling. In this case, the interaction between ICC_M and group size was also significant [$\Lambda = .9019, F_{(1,95808)} = 10423.71, p < .0001, \eta^2 = .0981$], indicating that the effects of these factors were especially pronounced in combination. Multilevel estimates of standard error of the mediated effect were essentially unbiased, whereas single-level estimates were often substantially downwardly biased. This bias could be greater than 20% of the true value of the standard error when ICC_M and group size were large. Such biases were attributable to an

appreciable downward bias in the single-level estimates of the standard error of the β_a effect, as well as more modest downward biases in the single-level estimates of the standard error of the β_b effect.

$2 \rightarrow 2 \rightarrow 1$. In the mediational chain in which a dichotomous group level predictor variable affected a group level mediator, the only possible modeling method for the first link (the β_a effect) was single-level (i.e., entirely group level) estimation. Estimates of the β_b effect, however, could be generated via either single-level or multilevel techniques, and using these results in the calculation of the mediated effect results in different estimates for the two techniques. Here, the conditions under which multilevel models showed the greatest advantage were those in which ICC_Y was relatively high [$\Lambda = .8053$, $F_{(1,95808)} = 23168.13$, $p < .0001$, $\eta^2 = .1947$] and those in which group sizes were large [$\Lambda = .8198$, $F_{(1,95808)} = 21056.50$, $p < .0001$, $\eta^2 = .1802$]. Under such conditions, the underestimation of single-level standard errors was on the order of 38 to 50%. Even under more ideal conditions with lower levels of ICC_Y and smaller group sizes, underestimation of the standard error was on the order of 10 to 15%. Since only single-level estimates were available for the β_a path, and these were essentially unbiased, the downward bias of the single-level estimates of the standard error of the mediated effect was entirely due to underestimation of the standard error of the β_b path.

Continuous Initial Variable Models

Three models were simulated in which a continuous initial variable began the mediational chain. The following section discusses results involving the relative bias in single-level and multilevel standard errors of the β_a estimate, the β_b estimate, and the mediated effect in each of the three model types.

$1 \rightarrow 1 \rightarrow 1$. For the mediational chain involving three individual level variables, the extent of single-level underestimation of the standard error of the mediated effect was a function of ICC_M [$\Lambda = .6212$, $F_{(1,95808)} = 58430.76$, $p < .0001$, $\eta^2 = .3788$], ICC_Y [$\Lambda = .8380$, $F_{(1,95808)} = 18519.94$, $p < .0001$, $\eta^2 = .1620$], and group size [$\Lambda = .5664$, $F_{(1,95808)} = 73347.38$, $p < .0001$, $\eta^2 = .4336$]. Interactions between ICC_M and group size [$\Lambda = .8917$, $F_{(1,95808)} = 11636.41$, $p < .0001$, $\eta^2 = .1083$] and between ICC_Y and group size [$\Lambda = .9493$, $F_{(1,95808)} = 5113.91$, $p < .0001$, $\eta^2 = .0507$] also affected single-level estimates, indicating that the combination of these effects had a particularly strong effect on relative bias values. Under the most extreme conditions (high ICC_M , high ICC_Y , and large group size), the downward bias in the single-level estimates of the standard error of the mediated effect was over 20% of the true value of the parameter. These biases could be attributed to appreciable downward bias in single-level estimates of the

standard error of β_a in those conditions with relatively large values of ICC_M , combined with more modest biases in single-level estimates of the standard error of β_b . In the most extreme conditions, the bias in the standard error of β_b was fairly substantial (19%) as well.

2 → 1 → 1. In the models in which a continuous group level predictor affected an individual level mediator, $ICC_M [\Lambda = .7700, F_{(1,95808)} = 28619.23, p < .0001, \eta^2 = .2300]$ and group size $[\Lambda = .7347, F_{(1,95808)} = 34598.88, p < .0001, \eta^2 = .2653]$ were the primary influences which determined the extent to which multilevel estimates of the standard error of the mediated effect improved upon single-level estimates. For small groups, the underestimation was on the order of 9 to 23%. For larger groups, the underestimation was on the order of 23 to 43%. This bias could be primarily attributed to substantial bias in the single-level OLS estimates of the standard error of the β_a effect, combined with more modest bias in the estimates of the standard error of β_b .

2 → 2 → 1. For the mediational chain in which both predictor and mediator were continuous group level variables, the primary determinants of the extent of multilevel advantage over single-level estimation were $ICC_Y [\Lambda = .8278, F_{(1,95808)} = 19926.09, p < .0001, \eta^2 = .1722]$ and group size $[\Lambda = .8416, F_{(1,95808)} = 18034.86, p < .0001, \eta^2 = .1584]$. For small groups, the underestimation of the standard error of the mediated effect in single-level models ranged from 10 to 28%, while for larger groups the underestimation ranged from 19 to 48%. This bias was entirely due to underestimation of the standard error of β_b .

Effects of β_a and β_b

In all models except 2 → 2 → 1, the true value of the β_b parameter interacted with analysis type as a significant predictor of the bias in the standard error of the mediated effect. In other words, the extent of the advantage of multilevel over single-level modeling depended on the true value of β_b . However, this factor was not a significant predictor of the biases in the standard errors of either β_a or β_b . Therefore, we can conclude that the effect of β_b on the bias of the standard error of the mediated effect arises from the equation used to calculate this standard error (Equation 5). In this equation, the estimate of the standard error of the β_a coefficient is squared and multiplied by the estimate of β_b squared. The effect, then, occurs because a larger β_b value weights any bias which may be present in β_a more heavily in calculating the standard error of the mediated effect. If the single-level estimate of the standard error of β_a is negatively biased, then more of this bias will appear in the single-level estimate of the standard error of the

mediated effect when β_b is large (as in the dichotomous $2 \rightarrow 1 \rightarrow 1$ model). If the single-level OLS estimate of the standard error of β_a is essentially unbiased, then more bias may be present in the single-level estimate of the standard error of the mediated effect when β_b is small (as in the dichotomous $1 \rightarrow 1 \rightarrow 1$ model). The effect sizes associated with the β_b variable ranged from $\eta^2 = .0571$ and $\eta^2 = .0728$ in the dichotomous $2 \rightarrow 1 \rightarrow 1$ and $1 \rightarrow 1 \rightarrow 1$ models to $\eta^2 = .1123$ and $\eta^2 = .1382$ in the continuous $1 \rightarrow 1 \rightarrow 1$ and $2 \rightarrow 1 \rightarrow 1$ models.

The effects of the true value of β_b on the relative bias in the standard error of the mediated effect, then, arise simply from the mathematical formula used to calculate the standard error of the mediated effect. We would expect β_a to operate similarly, simply weighting any bias in the standard error of β_b in the calculation of the standard error of the mediated effect. However, the effects of the true value of β_a , though significant at the traditional $p < .05$ level, did not achieve the $\eta^2 > .05$ criterion used in this study. This difference may be due to the larger range of true β_b values (.1 - .4, as opposed to .3 - .6 for the true β_a values) included in the simulation design.

Patterns of Simulation Results

In general, results of the simulation study showed downward biases associated with single-level estimation of standard errors of mediated effects in clustered data. This effect was present, if modestly so, even in $1 \rightarrow 1 \rightarrow 1$ models involving no group level predictors. Large group sizes, high ICCs of the mediator, and high ICCs of the outcome variable were associated with a larger degree of downward bias in estimates of the standard error of the mediated effect. Therefore, the advantage to be gained by using a multilevel mediational model, which showed no such bias, would be greatest under these conditions. The number of groups in the analysis had relatively little effect on the extent of the single-level bias and the multilevel advantage.

The magnitude of the bias in the single-level standard error estimates was especially large in certain models. In particular, the $2 \rightarrow 1 \rightarrow 1$ models and the $2 \rightarrow 2 \rightarrow 1$ models all had multiple conditions which showed downward biases of over 20%. Generally, these large biases in the standard error of the mediated effect were traceable to substantial downward biases in the $2 \rightarrow 1$ path. For the $2 \rightarrow 1 \rightarrow 1$ models, this was bias in the β_a path; for the $2 \rightarrow 2 \rightarrow 1$ models, this was bias in the β_b path. Models which include such a $2 \rightarrow 1$ link, then, seem to be particularly vulnerable to bias in single-level standard error estimates, and for such models, multilevel mediational modeling may provide much more accurate tests of the mediated effect.

The effects of the true value of β_b , as well as potential effects of β_a , on bias in the standard error of the mediated effect are entirely predictable, given the expression used to calculate the approximate standard error. For situations with strong relationships between the mediator and outcome, any bias in estimating the standard error of β_a will be heavily weighted in calculating the standard error of the mediated effect, resulting in greater bias than a situation with a weaker relationship. Thus the choice of an analytic method that minimizes bias (i.e., multilevel EB/ML estimation rather than single-level OLS estimation for clustered data) may be more important when the effects in the model are particularly strong, especially when the data exhibit other features associated with bias (e.g., large group size, high ICC_M or ICC_Y , or a $2 \rightarrow 1$ link in the mediational chain).

ATLAS Example

Data from the evaluation trial of the Adolescents Training and Learning to Avoid Steroids (ATLAS) program were analyzed in order to illustrate the differences between single-level and multilevel mediational modeling in a real-world data set. The ATLAS program is a team-based educational intervention program designed to decrease steroid use among high school football players. The program was designed to achieve its effects through changing a number of mediators believed to be related to steroid use, or, more proximally, to intentions to use steroids. These targeted mediators included knowledge constructs, peer and non-peer influence measures, attitudes, and individual player characteristics. The program also promoted nutrition behaviors and strength training self-efficacy as healthy alternatives to steroid use in improving athletic performance. Details of the prevention program can be found in Goldberg et al. (1996), and results of individual level mediational analyses can be found in MacKinnon et al. (2001).

Thirty-one schools were included in the ATLAS study, with 15 assigned to the experimental intervention and 16 to the non-intervention control condition. Data were collected from 17 to 66 players on each school's football team ($\bar{X} = 40$, $SD = 15$). Pretest measures were gathered prior to intervention, and posttest measures were obtained immediately following the seven-week intervention period.

Two sample analyses are presented below. The first of these examines the effects of perceived tolerance of steroid use by the individual player's (a) coach, (b) friends, (c) peers, and (d) parents as mediators of the relationship between the intervention program and intentions to use steroids. This case illustrates a series of $2 \rightarrow 1 \rightarrow 1$ models, with the team-level intervention program affecting individual perceptions, and, ultimately, individual

intentions to use steroids. The second analysis concerns individual and mean team levels of strength training self-efficacy as mediators of the relation between the intervention and intentions to use steroids. This example shows a $2 \rightarrow 1 \rightarrow 1$ model involving the individual level mediator in combination with a $2 \rightarrow 2 \rightarrow 1$ model involving the mean team effect as a mediator. Detailed information about the measures used in these sample analyses can be found in Goldberg et al. (1996).

Analyses were conducted by estimating single-level Equations 1-3 and their multilevel equivalents (from Table 1) to obtain single-level and multilevel estimates of $\hat{\beta}_a$ and $\hat{\beta}_b$, which were then combined to produce the single-level and multilevel $\hat{\beta}_a\hat{\beta}_b$ estimates of the mediated effect. Each equation included pretest measures of all included variables as covariates in order to isolate post-intervention changes in the variables. Single-level and multilevel standard error estimates were calculated using the multivariate delta method/first order Taylor series expression given in Equation 5.

ATLAS Results

Example 1: Individual Perceptions of Tolerance of Steroid Use

In the first example, the dichotomous group level intervention is believed to affect the individual player's perceptions of (a) the team coach's tolerance of steroid use, (b) the player's friends' tolerance of steroid use, (c) the player's peers' tolerance of steroid use, and (d) the player's parents' tolerance of steroid use. Changes in these mediators, then, would be expected to cause change in the individual's intentions to use steroids. Results of the single-level and multilevel mediational analyses are shown in the top section of Table 2. Both analyses indicated that only one of the four tolerance measures, perceived coach tolerance, did, in fact, significantly mediate the relationship between the intervention and intention to use steroids. Examination of the β_a and β_b paths that make up this mediated effect indicated that the ATLAS intervention decreased the level of perceived coach tolerance of steroid use, and that perceived level of coach tolerance was positively related to intentions to use steroids, so that the intervention-induced decreases in perceived coach tolerance resulted in decreased intentions to use. The single-level standard error of the mediated effect for this variable was about 29% smaller than the multilevel standard error. Looking at the standard errors of the two component paths, the single-level standard error of β_a was 43% smaller than the multilevel standard error, and essentially no difference was found for the β_b path. This pattern is quite similar to those found in the simulation study, where the single-level standard

Table 2

Single-level and Multilevel Estimates (and Standard Errors) of β_a , β_b , and the Mediated Effect for ATLAS Mediation Models

Mediator	Analysis Type					
	Single-Level			Multilevel		
	β_a	β_b	Mediated Effect	β_a	β_b	Mediated Effect
Example 1						
Perceived coach tolerance of steroid use	-.3734*** (.0717)	.0924*** (.0271)	-.0345** (.0121)	-.3629** (.1252)	.0915*** (.0272)	-.0332** (.0170)
Perceived friend tolerance of steroid use	-.1314 (.0982)	.0730*** (.0202)	-.0096 (.0076)	-.1207 (.1221)	.0722*** (.0202)	-.0087 (.0091)
Perceived peer tolerance of steroid use	-.1461 (.1000)	.0518** (.0200)	-.0076 (.0059)	-.1672 (.1560)	.0508* (.0201)	-.0085 (.0086)
Perceived parent tolerance of steroid use	-.1710* (.0745)	.0015 (.0258)	-.0003 (.0044)	-.1592 (.1044)	.0003 (.0258)	.0001 (.0041)
Example 2						
Individual strength training self-efficacy	.4068*** (.0586)	-.1634*** (.0293)	-.0666*** (.0153)	.4174*** (.0996)	-.1636*** (.0292)	-.0683*** (.0203)
Mean team strength training self-efficacy	.4850*** (.0984)	-.2250* (.1134)	-.1091 [†] (.0593)		-.2156 [†] (.1240)	-.1046 (.0638)

[†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

errors were often substantially underestimated when a higher level variable was predicting a lower level variable (i.e., along a 2 → 1 link). Here, substantial underestimation was present along the path where the team level intervention affected individual level perceptions of coach tolerance.

Both the single-level mediational analysis and the multilevel mediational analysis showed that neither the friend tolerance nor the peer tolerance measure were significant mediators of the effect of the ATLAS intervention on intentions to use steroids. Examination of the individual β_a and β_b paths showed that even though these two constructs were significantly related to intentions to use steroids (i.e., the β_b paths were significant), neither construct

was affected by the ATLAS intervention (i.e., the β_a paths were non-significant). This suggests a failure of action theory, rather than a failure of conceptual theory. Again, the smaller standard errors in the single-level analysis (16% smaller for perceived friend tolerance and 31% smaller for perceived peer tolerance) were due to smaller standard errors of the β_a path in which the group level intervention variable affected the individual level mediator. No such underestimation was found in the standard error of the β_b path involving the individual level mediator and the individual level outcome.

Both the single-level and the multilevel mediational analysis showed that perceived parental tolerance was not a significant mediator of the intervention effect on intentions to use steroids. The single-level estimate of the standard error of the β_a path in this analysis was 28% smaller than the multilevel estimate. With this smaller standard error, the single-level analysis would suggest that the ATLAS intervention did, in fact, decrease perceived parent tolerance. However, the multilevel analysis, with its more accurate standard error, showed that this decrease did not achieve conventional levels of significance. Moreover, both analyses showed that the perceived parent tolerance mediator was not related to intentions to use steroids.

Example 2: Individual and Team Levels of Strength Training Self-Efficacy

The second example involves both individual strength training self-efficacy and mean team levels of strength training self-efficacy as mediators of the intervention effect on intentions to use steroids. Here, the team mean on strength training self-efficacy may operate as a contextual effect: high mean levels reflect a context in which strength training is generally accepted as an effective means of improving athletic performance. The model incorporates both individual and team mean levels of the construct simultaneously, to provide conservative estimates of the mean effect on intentions, above and beyond the individual effect.

Results of these analyses are shown in the bottom section of Table 2. Both the single-level and the multilevel mediational model showed that individual strength training self-efficacy significantly mediated the relationship between the ATLAS intervention and intentions to use steroids. The intervention increased self-efficacy, and these increases led to decreases in intentions. The single-level standard error of this mediated effect was 25% smaller than the multilevel standard error, and this difference was due to underestimation (by 41%) of the standard error of the β_a path. Again, the underestimation occurred along a 2 → 1 path relating the group level intervention to the individual level mediator. In this case,

however, the inferences drawn about the relationships among the variables were similar for the two types of analyses, regardless of the underestimation.

Both single-level and multilevel mediational analyses also indicated that the contextual effect of team mean strength training self-efficacy was not a significant mediator of the intervention effect (though the single-level mediated effect would be significant at the .05 level in a one-tailed test). Because the β_a path involves only group level variables (i.e., group mean self-efficacy is predicted by group assignment to intervention or control conditions) a single-level model is appropriate. Examination of the β_a coefficient for the group level effect indicated that the ATLAS program did, in fact, increase team mean self-efficacy. Examination of the single-level β_b path indicated that these changes in the mean team strength training self-efficacy construct were associated with changes in intentions to use steroids, above and beyond the individual level effect. However, the multilevel β_b path, tested with a larger and more accurate standard error, was not significant at the $p < .05$ level, suggesting that this mean construct was not related to individual intentions.² The single-level estimate of the standard error of the team mediated effect was only 7% smaller than the multilevel estimate. However, this case illustrates that even a quite moderate level of underestimation could potentially result in different inferences based on single-level analyses in situations where multilevel analysis is actually more appropriate. Again, the single-level underestimation of the standard error of the mediated effect could be traced back to underestimation of the standard error of a $2 \rightarrow 1$ path. Here, the β_b path relating the group level mediator to the individual level outcome variable was underestimated by 8.5%.

General Discussion

The simulation study consistently showed underestimation of the standard error of the mediated effect in single-level models with clustered data. Group size, ICC of the mediator, and ICC of the outcome variable were identified as factors which increase the extent of such underestimation. Consequently, these factors also define the conditions under which multilevel mediational modeling will be most advantageous. Results of the simulation study also illustrate how actual values of β_a and β_b can affect the extent of single-level underestimation through their role in the calculation of the standard error of the mediated effect. Single-level underestimation of the

²The failure to detect this relationship in the multilevel analysis should, of course, not be viewed as definitive evidence that the effect does not exist. It is possible that with the sample size in this dataset, the analysis did not have sufficient power to detect the contextual effect.

mediated effect could often be traced back to underestimation along a $2 \rightarrow 1$ link (i.e., underestimation of the standard error of β_a in a $2 \rightarrow 1 \rightarrow 1$ model and underestimation of the standard error of β_b in a $2 \rightarrow 2 \rightarrow 1$ model). Results of the simulation study also suggest that the level 2 sample size (i.e., the number of groups in the analysis) plays at most a minimal role in determining the underestimation of single-level standard errors and the multilevel advantage in this regard. However, this variable is known to play a key role in other aspects of multilevel model performance, most notably power (Kenny, Kashy, & Bolger, 1998; Murray, 1998).

The primary findings from the simulation study were borne out by the ATLAS examples. In particular, the single-level standard errors of mediated effects in both examples were smaller than the corresponding multilevel standard errors. We may infer from the simulation study that these smaller standard errors were, in fact, the result of underestimation in the single-level model. Moreover, as in the simulation study, the underestimation of the standard errors of mediated effects in the ATLAS examples could typically be traced back to underestimation along a $2 \rightarrow 1$ link. The extent of the underestimation was smaller in the ATLAS examples than in the simulation study. This may be due, in part, to the more complex multiple mediator models tested with the ATLAS data. Differences in the properties of the real and simulated variables may also be partially responsible. For example, the ATLAS analysis involving mean team level of strength training self-efficacy showed only a 7% difference between single-level and multilevel standard errors of the mediated effect. In contrast, underestimation was on the order of 35 to 50% in the parallel $2 \rightarrow 2 \rightarrow 1$ simulation model. However, the residual ICC of the ATLAS outcome variable was quite small ($ICC_Y = .005$) when both individual and team levels of strength training self-efficacy were included as predictors in the model. This value was only a tenth of the size of the smallest residual ICC value included in the simulation. Nevertheless, the example illustrated that single-level and multilevel analyses may result in different inferences even with such a small degree of within-group homogeneity.

Each of the multilevel equations presented in Table 1 and applied in the simulation study and the ATLAS example includes only a single random coefficient, the intercept term. This allows each group intercept, representing an overall group level of the dependent measure, to vary across the set of groups. The error term associated with this random intercept enables the modeling of the correlated error inherent in clustered data. All other coefficients in the lower level equations are treated as fixed. However, in $1 \rightarrow 1 \rightarrow 1$ and $2 \rightarrow 1 \rightarrow 1$ models (i.e., whenever the mediator is a level 1 variable), it would also be possible to specify the β_b term as random. This would involve the specification of an

additional level 2 model in which a group's relationship between mediator and outcome is predicted from an overall level of relationship and an additional group level error term. In this simulation study and in the ATLAS example, the β_b coefficient was always treated as fixed, to permit comparisons across model types and to fixed single-level effects. In some cases, however, specification of β_b as random may be appropriate, particularly if theory dictates considerable variability in the strength of the mediator-outcome relationship or if the groups in the study are very heterogeneous. Specifying an additional coefficient as random, however, increases the computational demand of the solution, and the addition can affect both estimates of fixed effects and standard errors in a rather complex fashion (see Kreft & deLeeuw, 1998, p. 126-130). Moreover, the addition of a random slope will make decisions about the centering of lower level variables more critical. With a random slope, centering decisions will affect the variability of intercepts across groups, as well as the covariance of the random intercepts with the random slopes. Centering may also provide a computational advantage, as the correlation between random intercepts and slopes is often reduced (Kreft, de Leeuw, & Aiken, 1995). Hoffmann and Gavin (1998) provide a thoughtful discussion of the possible centering strategies that could be used in examining multilevel mediation, as well as other types of research paradigms.

The procedures of multilevel mediational analysis, presented here in the context of individuals nested within groups, are equally applicable to repeated measures nested within persons. Kenny, Kashy, and Bolger (1998) briefly discuss two situations, termed "lower level mediation" and "upper level mediation", which parallel in a repeated measures context the $1 \rightarrow 1 \rightarrow 1$ and $2 \rightarrow 1 \rightarrow 1$ models presented here. More complex longitudinal mediational models are also testable using extensions of multilevel and latent growth modeling techniques (e.g., Cheong, MacKinnon, & Khoo, 1998).

Many of the difficulties in the interpretation of single-level mediational models apply equally to and may be complicated by the structure of multilevel mediational models. Holland (1988) has outlined the difficulty of making causal statements in the analysis of mediational models even when the independent variable represents random assignment to conditions, as in the ATLAS examples. The existence of multiple levels of analysis further complicates interpretation. It would seem that the group-level program effect estimate on a group-level outcome variable and the group-level program effect on the mediator are generally accurate estimates of causal effects given true random assignment of groups to conditions. It is less clear whether causal interpretations are available for the interpretation of effects on individuals when random assignment occurs at the group level. A detailed study of the underlying causal relationships in multilevel models using Rubin's causal model (Rubin, 1974) is beyond the scope of this article. It is clear that the relationship

between the mediator and the dependent variable is not necessarily an accurate estimate of the true relationship between the mediator and the outcome. For example, the relationship between the mediator and the outcome may differ across the groups or the direction of the relationship may be inaccurate where the outcome actually causes the mediator (MacCallum, 1986; Spirtes, Glymour, & Schienens, 1993). Given the ambiguity in interpretation of mediational models, whether single-level or multilevel, it is best to incorporate a variety of additional research strategies to clarify mediational processes. These additional research strategies include replication studies with randomization to conditions based on mediational processes found to be important in prior studies (West & Aiken, 1997). Meta-analyses of critical relationships are also important (Cook et al., 1992), as are theoretical arguments for the specificity of effects through one mediator and not others.

References

- Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, 6, 267-285.
- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, 8, 158-233.
- Burstein, L. (1985). Units of analysis. *International encyclopedia of education*. London: Pergamon Press.
- Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Cheong, J., MacKinnon, D. P., & Khoo, S. (1998, October). *A latent growth modeling approach to mediation analysis*. Poster presented at the conference on New Methods for Analysis of Change. State College, PA.
- Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1992). *Meta-analysis for explanation*. New York: Sage.
- de Leeuw, J. (1992). Series editor's introduction to hierarchical linear models. In A. S. Bryk & S. W. Raudenbush, *Hierarchical linear models: Applications and data analysis methods* (pp. xiii-xvi). Newbury Park, CA: Sage.
- Goldberg, L., Elliot, D., Clarke, G. N., MacKinnon, D. P., Moe, E., Zoref, E. L., Green, C., Wolf, S. L., Greffrath, E., Miller, D. J., & Lapin, A. (1996). Effects of a multidimensional anabolic steroid prevention intervention: The Adolescents Training and Learning to Avoid Steroids (ATLAS) program. *Journal of the American Medical Association*, 20, 1555-1562.
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55, 708-713.
- Hansen, W. B. (1992). School-based substance abuse prevention: A review of the state of the art in curriculum, 1980-1990. *Health Education Research*, 7, 403-430.

- Hauck, W. W. & Anderson, S. (1984). A survey regarding the reporting of simulation studies. *American Statistician*, 38, 214-216.
- Hedeker, D., Gibbons, R. D., & Flay, B. R. (1994). Random effects regression models for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, 62, 757-765.
- Hofmann, D. A. & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations, *Journal of Management*, 24, 623-641.
- Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 449-493). Washington, DC: American Sociological Association.
- Judd, C. M. & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602-619.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. I, pp. 233-265). New York: McGraw-Hill.
- Kreft, I. G. G. (1997). The interactive effect of alcohol prevention programs in high school classes: An illustration of item homogeneity scaling and multilevel analysis techniques. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 251-277). Washington, DC: American Psychological Association.
- Kreft, I. & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.
- Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30, 1-22.
- Krull, J. L. & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23, 418-444.
- MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, 100, 107-120.
- MacKinnon, D. P. & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17, 144-158.
- MacKinnon, D. P., Goldberg, L. Clarke, G. N., Elliot, D. L., Cheong, J., Lapin, A., Moe, E. L., & Krull, J. L. (2001). Mediating mechanisms in a program to reduce intentions to use anabolic steroids and improve exercise self-efficacy and dietary behavior. *Prevention Science*, 2, 15-28.
- MacKinnon, D. P., Johnson, C. A., Pentz, M. A., Dwyer, J. H., Hansen, W. B., Flay, B. R., & Wang, E. Y. (1991). Mediating mechanisms in a school-based drug prevention program: First-year effects of the Midwestern Prevention Project. *Health Psychology*, 10, 164-172.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41-62.
- Mood, A., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. New York: McGraw-Hill.
- Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics*, 32, 385-397.
- Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.
- Palmer, R. F., Graham, J. W., White, E. L., & Hansen, W. B. (1998). Applying multilevel analytic strategies in adolescent substance use prevention research. *Preventive Medicine*, 27, 328-336.
- Pedhazur, E. J. (1982). *Multiple regression is behavioral research: Explanation and prediction*. Fort Worth, TX: Holt, Rinehart, & Winston.

- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351-357.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, *66*, 688-701.
- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, *227*, 918-924.
- Scariano, S. M. & Davenport, J. M. (1987). The effects of violations of independence assumptions in one-way ANOVA. *The American Statistician*, *41*, 123-129.
- Scott, A. J. & Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *Journal of the American Statistical Association*, *77*, 848-854.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290-312). Washington, DC: American Sociological Association.
- Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), *Sociological methodology 1986* (pp. 159-186). Washington, DC: American Sociological Association.
- Spirtes, P., Glymour, P., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.
- Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research*. New York: Macmillan.
- Walsh, J. E. (1947). Concerning the effect of the intraclass correlation on certain significance tests. *Annals of Mathematical Statistics*, *18*, 88-96.
- West, S. G. & Aiken, L. S. (1997). Toward understanding individual effects in multi-component prevention programs: Design and analysis strategies. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 251-277). Washington, DC: American Psychological Association.

Copyright of Multivariate Behavioral Research is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.