Journal of Personality and Social Psychology
2002, Vol. 83, No. 1, 126–137

0022-3514/02/$5.00 DOI: 10.1037//0022-3514.83.1.126

The Statistical Analysis of Data From Small Groups

David A. Kenny
University of Connecticut

Lucia Mannetti, Antonio Pierro, and Stefano Livi
University of Rome Sapienza

Deborah A. Kashy
Michigan State University

The authors elaborate the complications and the opportunities inherent in the statistical analysis of small-group data. They begin by discussing nonindependence of group members' scores and then consider standard methods for the analysis of small-group data and determine that these methods do not take into account this nonindependence. A new method is proposed that uses multilevel modeling and allows for negative nonindependence and mutual influence. Finally, the complications of interactions, different group sizes, and differential effects are considered. The authors strongly urge that the analysis model of data from small-group studies should mirror the psychological processes that generate those data.

Research designs involving small groups occur in many of the subdisciplines of psychology. For example, in clinical psychology, group psychotherapy outcome research necessarily involves small groups. Similarly, small groups are typically used in research in industrial–organizational psychology investigating team productivity. Social psychologists use small groups to study a variety of group processes, including interpersonal influence, group identification, and intergroup relations. Unfortunately, all too often, the analysis of group data is based on models that were developed for the analysis of data from individuals and not groups. For significant theoretical advances in the study of group processes to occur, it is essential that methods be developed that are specifically designed to model and analyze group data. The major goals of this article are to describe the problems inherent in modeling group data using models developed for individuals and to develop procedures that probe small-group data much more extensively than do those usually used by small-group researchers.

Several previous articles have discussed the analysis of data from small-group research and related topics (Anderson & Ager, 1978; Crits-Christoph & Mintz, 1990; Firebaugh, 1978; Glick, 1980; Griffin & Gonzalez, 1995; Haney, 1980; Hoyle, Georgesen, & Webster, 2001; Kenny, Kashy, & Bolger, 1998; Kenny & La Voie, 1985; Klein, Dansereau, & Hall, 1994). Work on this subject has been done not only by social psychologists but also by industrial–organizational, educational, and clinical psychologists, political scientists, and sociologists. There is much to be learned from these articles, and we reiterate many of their points. However, one major emphasis of this article is to advance a data-analytic model that is informed by social–psychological theory concerning the ways group members affect and influence one another. Moreover, we take full advantage of the newly developed statistical technique of multilevel modeling. Although this article is highly detailed in some sections, we should make clear that we in no way claim to exhaust the complete analysis of this topic. For instance, we do not consider sequential features of group interaction (e.g., Bakeman & Gottman, 1997) that need to be addressed if we are to have a complete treatment of group effects.

This article is divided into three parts. In the first part we consider the nonindependence of observations in small-group data, describing how nonindependence can be measured and how it affects research conclusions when it is ignored, as it all too often is. In the second section, we describe several analysis strategies that have been previously applied to small-group data, and we present a new model that allows researchers to examine simultaneously the effects of variables as well as how, within a group, group members affect one another. In the final section, we discuss several more complicated versions of the proposed model as well as limitations of the approach.

Nonindependence of Observations in Groups

Very often, the data from small-group studies are said to be *nonindependent,* which means that persons who are in the same group are more similar (or dissimilar) to one another than are persons who are members of different groups. Nonindependence can be viewed as a correlation between observations that may be either positive (i.e., two members from the same group are more similar than are two members from different groups) or negative (i.e., two members from the same group are more dissimilar than are two members from different groups). Nonindependence between the scores of group members undermines the statistical assumption (as in analysis of variance [ANOVA] and regression

David A. Kenny, Department of Psychology, University of Connecticut; Lucia Mannetti, Antonio Pierro, and Stefano Livi, Department of Psychology, University of Rome Sapienza, Rome, Italy; Deborah A. Kashy, Department of Psychology, Michigan State University.

This research was supported in part by National Institute of Mental Health Grant RO1-MH51964 to David A. Kenny. We thank Antonius Cillessen, who provided us with comments on a draft of this article.

Correspondence concerning this article should be addressed to David A. Kenny, Department of Psychology, University of Connecticut, Storrs, Connecticut 06269-1020. E-mail: kenny@uconnvm.uconn.edu

models) that observations are independent replications. There are not, then, as many degrees of freedom as might be thought, and there must be an adjustment in the analysis.

Sources of Nonindependence

Kenny and Judd (1986) described three factors that might produce nonindependence in groups: compositional effects, common fate, and mutual influence. A *compositional effect* occurs when people are nonrandomly sorted into groups. For example, it is well documented that marriage partners are very similar on a host of variables (e.g., age, education, and intelligence) even before they have ever met, a phenomenon called *assortative mating*. *Common fate* creates nonindependence within groups when members of groups coexist in the same environment. Through the process of common fate, people can be in a group even if they never meet if they are linked to the same person (e.g., clients receiving psychotherapy from the same therapist).

The third source of nonindependence, *mutual influence*, is arguably the most important factor that leads to nonindependence. As Kenny (1996) has suggested, mutual influence can be direct or indirect. In direct influence, there are reciprocal paths from the scores of individuals to each other. For example, the more positively one person in the group talks, the more the others in the group also talk positively. In indirect social influence, one aspect of a group member influences another aspect of other group members. For example, Campbell, Simpson, Kashy, and Rholes (2001) have shown that when individuals have highly avoidant attachment styles, their dating partners tend to respond more negatively to them in stressful situations. Note that influence can be negative, causing what has been called a *boomerang effect*: One person may influence another to be different from the person. Our focus in this article is to provide a method that can be used to assess the impact of indirect mutual influence in small groups.

The Measurement of Nonindependence

The intraclass correlation coefficient can be used to estimate the degree of nonindependence in small group studies, and there are two methods for estimating this correlation. The first approach uses ANOVA. When one estimates the intraclass correlation using ANOVA, group is treated as the independent variable in a one-way, between-subjects ANOVA. From this analysis, we obtain a mean square between groups, or MS_{BG}, and a mean square within groups, or MS_{WG}. With these two mean squares, we estimate the intraclass correlation, or r_I:

$$r_I = \frac{MS_{BG} - MS_{WG}}{MS_{BG} + (n - 1)MS_{WG}},$$

where n is the group size and is assumed to be equal across groups. If group sizes are unequal, the value of n must be computed using a formula presented in Kenny and La Voie (1985). The limits of this correlation are 1 and $-1/(n - 1)$. Notice that if the group size is two, then the intraclass correlation varies from 1 to -1; however, if the group size is greater than two, then the lower limit is greater than -1. In a triad[1] the lower limit of the intraclass correlation is $-.5$: If Person A is very different from both Person B and Person C, then B and C cannot also be very different from each other. The intraclass correlation equals 1 when all the scores in the group equal the same value and the group means differ ($MS_{BG} > 0$ and $MS_{WG} = 0$); it is at its minimum when the group means are equal but the scores within one or more groups vary ($MS_{WG} > 0$ and $MS_{BG} = 0$).

Tests of the intraclass correlation for statistical significance also use the mean squares from the ANOVA that treats group as the independent variable. The test is an F test computed as MS_{BG}/MS_{WG} with $k - 1$ degrees of freedom in the numerator and $k(n - 1)$ degrees of freedom in the denominator, where k is the number of groups and n is the group size. However, if the intraclass correlation were negative, then this F would be less than 1. Thus, if the intraclass correlation is negative, the F test should be computed as MS_{WG}/MS_{BG}, with a corresponding flip in the degrees of freedom. Therefore, unlike the usual F test in ANOVA, the test should be two-tailed, and the usual p value should be doubled.

The second and older approach to computing the intraclass correlation is a correlational method that has been revitalized by Griffin and Gonzalez (1995). Using the correlational method for group data (i.e., when group size is greater than two) is conceptually straightforward but operationally somewhat awkward. The procedure involves first creating every possible pair of scores in the group. So if there were four people in a group and the scores were 4, 6, 8, and 9, there would be 12 pairs of scores: {4, 6}, {4, 8}, {4, 9}, {6, 4}, {6, 8}, {6, 9}, {8, 4}, {8, 6}, {8, 9}, {9, 4}, {9, 6}, and {9, 8}. In each pair, the first score is referred to as the X variable, and the second is the Y variable. The intraclass correlation is then estimated on the basis of these pairs of scores for each of the groups using formulas for the standard Pearson correlation. As suggested by Griffin and Gonzalez (1995), a Z test can be used to evaluate whether the intraclass correlation differs significantly from zero. In this test, one computes Z by multiplying the intraclass correlation by the square root of the total number of participants in the study. The ANOVA and correlational measures of the intraclass correlation yield very similar estimates as long as there are five or more groups. Both measures have a slight negative bias (i.e., they tend to underestimate the level of nonindependence), but the bias from the correlational method is somewhat greater than that from the ANOVA approach.

Negative Nonindependence

Researchers and methodologists sometimes fail to recognize that nonindependence can be negative as well as positive. This problem is exacerbated by the fact that the intraclass correlation is sometimes interpreted as the proportion of variance explained by the group effect, and neither proportions nor variances can be negative. In spite of this interpretation, it is the case that the intraclass correlation is a correlation, and correlations (as well as nonindependence in general) can be negative as well positive.

What leads to a negative correlation between members of small groups? Group processes very often lead to differentiation as well as homogeneity. In particular, there may be a fixed resource such as time, status, power, or leadership that is the same or essentially the same for all groups. For instance, the percentage of time each individual spends talking in a group discussion (or some variable

[1] So if the group size is four, the lower limit is $-.33$, and if the group size is five, the lower limit is $-.25$.

that correlates with percentage of time spent talking, e.g., number of interruptions) might well have a negative intraclass correlation. Another example in which a negative intraclass correlation would emerge is when groups are given a fixed award (e.g., 10 euros) and are told to divide it among group members; the number of euros received by each member would have a negative intraclass correlation. Thus, if the environment or the experimenter structures the situation so that there is a fixed resource, a negative intraclass correlation might result.

Social comparison processes very often lead to negative intraclass correlations. For instance, in sporting events (e.g., a tennis match), one person's happiness generated by winning might be associated with the other person's unhappiness generated by losing. Sometimes the way that questions are worded may exacerbate social comparison processes, as would be the case if participants were asked how well they did relative to other group members.

Effect of Nonindependence on p *Values*

Why is nonindependence important in statistical testing? The basic statistical model used by ANOVA and multiple regression assumes that scores are independent. If the scores from two individuals were not independent, then there would be complications in the statistical analysis. It is generally acknowledged (Kenny et al., 1998) that violating the independence assumption is much more serious than violating the other standard assumptions of ANOVA and regression (i.e., normality and homogeneity of variance). Unfortunately, many small-group researchers still continue to ignore the problem of nonindependence (Hoyle et al., 2001).

Nonindependence does not bias the effect estimates of the independent variables. So mean differences between conditions or regression coefficients estimating the effects of a predictor variable on an outcome measure are unbiased even if data are analyzed using the individual as the unit when there is nonindependence. However, nonindependence does distort the estimate of the error variance, so standard errors, p values, confidence intervals, and most effect-size measures are invalid. It is notable that the biasing effects of nonindependence can result in either an increase in Type I errors or an increase in Type II errors, with a corresponding decrease in power. Kenny et al. (1998) have shown that the most important factor in making the test liberal or conservative is the degree of nonindependence: The greater the degree of nonindependence is, the greater the bias in the p values is. Additionally, group size is an important factor, such that for any specific level of nonindependence, the larger the group is, the greater the bias is.

There is a further complication in understanding the effects of nonindependence on significance testing, and it has to do with the type of the independent variable. An independent variable can vary between groups (e.g., some groups are composed of only men, and other groups are composed of only women) or within groups (e.g., there are two men and two women in each group). The defining feature of a within-group independent variable is that its mean is the same for every group. For groups composed of two men and two women, if gender were coded as 1 and -1, the mean value of gender would equal zero for every group. For a between-groups independent variable, the effect of a positive intraclass correlation is to make the test too liberal, and the effect of a negative correlation is to make the test too conservative. Just the opposite happens with a within-group independent variable. If the correlation is positive, the test of the effects of the independent variable is too conservative, and if the correlation is negative, that test is too liberal.

There is yet an additional complication. The independent variable may be neither entirely between groups nor entirely within groups. Again, if we consider gender as the independent variable, some groups may be all women, others all men, and others a mixture of women and men. In this case, gender would not be entirely between or within groups. It can be called a *mixed variable*. Another example of a mixed independent variable is identification with the group. Some individuals may strongly identify with their group, whereas other individuals show little identification with the group. In addition, some groups, on average, may be high in identification, and others may be low. Thus, identification varies both within groups and between groups.

What is the effect of nonindependence when one is testing the effects of a mixed independent variable and treating individual rather than group as the unit of analysis? Recall that with a positive intraclass correlation, if the independent variable is between groups, tests tend to be overly liberal, and if the independent variable is within groups, tests tend to be overly conservative. Because a mixed variable varies both between and within groups, the test is pulled both by the liberalizing effects of its between-groups nature and by the conservative effects of its within-group nature, so there is relatively less bias in the p values. Thus, tests of mixed independent variables that treat the individual as the unit of analysis, even in the presence of nonindependence, may have little or no bias.[2]

Because the direction and degree of bias with a mixed variable depend on the degree to which the variable more strongly resembles a between-groups variable or a within-group variable, it can be helpful to index how similar it is to these two types of variables. An intraclass correlation can be computed for the independent variable, and the direction of that correlation indicates whether the variable is more similar to a between- or within-group variable. To compute the intraclass correlation, we dummy code the independent variable (if it is categorical in nature) and compute its intraclass correlation using one of the two approaches discussed earlier. If the estimated intraclass correlation is 1, then the independent variable is purely between groups; if the intraclass correlation is $-1/(n - 1)$, then the independent variable is purely within group. A mixed independent variable has some value between 1 and $-1/(n - 1)$. If the intraclass correlation is near 1, then the mixed independent variable acts more like a between-groups variable, and if the intraclass correlation is near $-1/(n - 1)$, the mixed independent variable acts more like a within-group variable. If the intraclass correlation for the independent variable is near zero, then the between- and within-group aspects of the variable are relatively balanced, creating little or no bias in the standard error or in tests that treat individual as the unit of analysis.

Consequential Nonindependence

Kenny et al. (1998) introduced the concept of *consequential nonindependence*, which refers to the smallest amount of noninde-

[2] As discussed in Kenny et al. (1998), nonindependence also lowers the degrees of freedom. However, the bias in significance testing due to a loss of degrees of freedom is ordinarily small.

pendence (i.e., intraclass correlation) that can lead to an unacceptably large p value (unacceptable being .10). So consequential nonindependence is the value of the intraclass correlation for which an inferential test yielding a p value of .05 really corresponds to a Type I error rate (i.e., alpha) of .10 or greater. The value of consequential nonindependence is surprisingly low—especially with large groups. For jury research ($n = 12$), the consequential nonindependence is only .040. This problem is exacerbated by the fact that once the group size is four or more, there is generally insufficient power to test for nonindependence if there are fewer than 25–30 groups in the study. For the jury research example, 408 jurors (34 groups) would be required for one to attain an adequately powerful test of whether the intraclass correlation differs significantly from zero. Thus, given the usual sample sizes used in research with groups of four or more, the test of the intraclass correlation may not be statistically significant, even though the actual intraclass correlation is large enough to bias the p value. As Kenny et al. (1998) stated, generally for small-group studies with four or more persons, the correct strategy is to assume that the data are nonindependent and not to bother to test whether the nonindependence is statistically significant.

Although we have focused this section on the consequences of ignoring nonindependence for statistical hypothesis testing, ignoring nonindependence can be much more serious in terms of its implications for progress in behavioral research. That is, when interdependence of individuals within groups is ignored, much of what is unique psychologically about interacting or working in groups is lost. Consider, for example, group psychotherapy outcome research. If a researcher focuses on individuals and ignores nonindependence, he or she may examine only how each individual's commitment to group therapy affects that individual's own outcomes from therapy. An important theoretical issue in group therapy concerns the impact that other group members have on an individual's outcomes. Thus, it may be that a group therapy client can benefit from therapy if the other members of the group are highly committed to therapy, perhaps regardless of the individual's own level of commitment. The point is that not only does nonindependence bias hypothesis tests when it is ignored, but ignoring it also reduces the quality of small-group research by limiting the kinds of questions such research can address. In the next section of this article, we detail several analysis strategies that allow researchers to conduct hypothesis tests that are not biased by nonindependence and that can delve into questions of how individuals within groups affect one another.

Analysis Strategies

Because there are standard methods for analyzing the effects of purely between-groups and purely within-group independent variables that take nonindependence into account (Kenny et al., 1998), our focus in this section is on a series of analysis strategies that specifically assess the effects of mixed independent variables. One of the unique features of mixed independent variables is that they allow the researcher to study their effects both at the level of the individual and at the level of the group. We begin with analyses that may be more familiar to the reader. The limitations of these methods serve as a way to introduce the new analytic approach to small-group data developed in this article. In our discussion, we focus on how each data-analytic method treats the issue of nonindependence. We also try to develop analysis strategies that make sense in light of social–psychological theory concerning the impact that individuals within groups have on one another.

We illustrate the different analysis strategies using data collected by Pierro and Livi (2000). In this data set there are eight 8-person groups in which all participants were women with an average age of 20.5. The task was a cooperative group problem-solving task, and the outcome that we consider is the individual's satisfaction with her own performance, measured on a 7-point scale. The exact wording in English is, "Are you satisfied with your performance in the group?" The independent variable is the number of acts emitted and received by a group member during the 40-min interactions. Acts were coded independently by two raters using Bales's (1950) interaction process analysis, and the total represents the average of the two raters. The range for the number of acts is 6–377.5, the mean is 97.0, and the standard deviation is 77.5. To make the coefficients from our analyses easier to interpret, we divided the number of acts by 100, so effects refer to hundreds of acts. Throughout this section, we denote the independent variable for an individual, number of acts, as X and the outcome variable for that individual, satisfaction, as Y. The mean of X for a group (the average number of acts computed across all group members) is denoted as M_X, and the mean of X not including the person (the average number of acts computed across all group members except the individual under consideration) is denoted as M'_X. There are N persons ($N = 64$ in the example) who are members of k groups ($k = 8$) of size n ($n = 8$), so $N = nk$.

The Standard Two-Step Method

The strategy that is generally recommended for studying group effects (see, e.g., Myers, 1979) involves two steps. The first step is to measure and test the intraclass correlation for the dependent measure. If that correlation is statistically significant, usually according to a liberal test of alpha equal to .25, then there is evidence of nonindependence. Given this circumstance, the standard recommendation is to make group the unit of analysis by using the group means of X and Y. Using the example data set, we have $MS_{BG} = 0.194$, $MS_{WG} = 1.792$, and $n = 8$. The intraclass correlation for the outcome variable, satisfaction, equals $-.125$ [(0.194 − 1.792)/(0.194 + 1.792(7))]. The test that the intraclass correlation is different from zero is statistically significant, $F(56, 7) = 9.24$, $p = .0048$. The negative correlation implies that if someone is very satisfied, the other group members are dissatisfied. The correlation is very near its theoretical lower limit of $-.143$.

Because of the statistically significant intraclass correlation, we compute group means and conduct the analysis using them. In the example, averages for number of acts and satisfaction are computed. The group averages are then used in the analysis such that the group average on X is used to predict the group average on Y. We find that the regression coefficient is .126, $F(1, 6) = 0.89$, $p = .382$. Using this analysis, we would have concluded that acts did not affect satisfaction.

This two-step strategy is problematic, however, when X varies both between and within groups (i.e., X is mixed), because when the researcher computes a group mean, much of the variation in X is lost. Thus, the use of the group mean approach to control for nonindependence is not workable when X is mixed. For the example, the number of acts, X, varies within groups as well as between groups. The intraclass correlation for X is .291 (not all that close to 1.000), and within-group variance is ignored in this analysis.

Note that if the variation in X were entirely between groups, group members' scores on X would all be the same and the group-mean approach would be a viable solution. Even in this case, however, we recommend the approach that we develop later because it more optimally weights the group means when group sizes vary.

Naive Analysis

By far the most common approach to group data is to do separate analyses for individuals and groups. In the individual-as-unit analysis, for each individual, Y is regressed on X. This analysis estimates the degree to which a person who scores higher in the number of acts also scores higher on satisfaction. In the group-as-unit analysis, M_Y is regressed on M_X, estimating (as described above) whether groups that are higher in the number of acts are also higher in satisfaction. The individual-as-unit analysis is a very common approach to analyzing group data, whereas it is somewhat less common for researchers to conduct an additional analysis using the group means. Researchers often assert, "Because I am only interested in individuals, I did an analysis with individual as the unit." However, such an approach is indefensible because it violates the independence assumption. Even if the substantive interest focuses on the behavior of individuals, the violation of the independence assumption typically invalidates significance testing results.

Moreover, the two analyses are partially redundant. Because group effects are contained in the individual scores, the individual analysis is not really purely at the individual level (Glick & Roberts, 1984). To see this, consider the extreme situation in which every person had exactly the same score on the independent and the outcome variables. In such a case, the regression coefficients (or mean differences if the independent variable were categorical) would be identical in the two analyses.

For the example, we previously reported the results of the analysis using group means. The analysis using individual as the unit shows that the regression coefficient using acts to predict satisfaction is .761 and is statistically significant, $F(1, 63) = 17.02$, $p < .001$. Thus, individuals with more acts were more satisfied.

Contextual Analysis

For many years, the only analysis model explicitly for group data was a method that came to be known as *contextual analysis* (Boyd & Iverson, 1979). This approach was used primarily in sociology and political science and was largely ignored within psychology. The method uses both X and M_X to predict Y. In essence, contextual analysis is like a naive analysis, but it combines the two analyses into a single analysis. The individual and the group compete to explain the variation in the outcome variable. For the example, this analysis tests whether the number of acts by the individual rather than the average of the group members better predicts the individual's level of satisfaction. We find that the effect of the individual is $B = 1.025$, $F(1, 62) = 23.33$, $p < .001$, whereas the effect of the group is $B = -0.900$, $F(1, 62) = 5.28$, $p = .025$. The effects, then, are in the opposite direction, and both are statistically significant.

The major problem with contextual analysis is that it likely violates the assumption of independence. The analysis presumes that all of the nonindependence is modeled because M_X is included in the analysis. In essence, this approach presumes that the nonindependence is entirely compositional and that it is entirely explained by M_X. Thus, this approach does not allow for nonindependence that is a result of mutual influence.

Between–Within Analysis

The prior two methods have statistical problems because they violate the independence assumption. A third approach, known as a between–within analysis, is not subject to this statistical problem and so offers an alternative strategy. If we examine the formula for the intraclass correlation, we see that it partitions the variance in the outcome variable into two sources: between and within groups. We have also noted that the variation of a mixed independent variable is composed of both between- and within-group variation. In essence, a between–within analysis uses the between-groups variation in the independent variable to predict the between-groups variation in the outcome measure. It also uses the within-group variation in the independent variable to predict the within-group variation in the outcome measure.

As we have just implied, in a between–within analysis, two analyses are done. In the between-groups analysis, the group means are analyzed, weighted by group size. Thus, the average number of acts is used to predict the average satisfaction. There are $k - 2$ degrees of freedom in the between-groups analysis, where k is the number of groups.

For the within-group analysis, each score has the group mean subtracted. Thus, $X - M_X$ is computed for each individual, as is $Y - M_Y$. Then the deviation score for the independent variable is used to predict the deviation score on the outcome measure. For the example, we test whether a person who is relatively high in the number of acts (relative to members of her own group) is also relatively high in personal satisfaction. The degrees of freedom for the within analysis are $k(n - 1) - 1$. There are some statistical complications for the within-group analysis. A set of $k - 1$ dummy codes needs to be created to indicate group membership, and these variables should be included in the within-group analysis to make sure that the degrees of freedom are correct, even though these group dummies do not explain any of the variance. With the between–within approach, the effect for the between-groups analysis is $B = 0.126$, $F(1, 6) = 0.89$, $p = .382$. In the within-group analysis, we find the effect is $B = 1.025$, $F(1, 55) = 21.38$, $p < .001$. Thus, the within-group effect is statistically significant, and the between-groups effect is not.

Even though this analysis method represents a statistically adequate solution to nonindependence, the approach does suffer from the problem that it is difficult to attach conceptual meaning to the two analyses, particularly the within-group analysis. Moreover, there is no direct test of differences between the effects of the two analyses.[3] Likely for these reasons, the between–within approach

[3] Below, we discuss a multilevel analysis that is equivalent to a between–within analysis. When we conduct that analysis, the test that the between-groups and within-group coefficients are equal fails ($p = .001$). Thus, the between-groups and the within-group coefficients are in fact statistically different.

has been infrequently used (however, see Fletcher and Fitness, 1990).

Actor–Partner Interdependence Model

In a series of articles, several authors (Kashy & Kenny, 2000; Kashy & Snyder, 1995; Kenny, 1996; Kenny & Cook, 1999; Kenny et al., 1998) have proposed a model that they have called the actor–partner interdependence model (APIM). This model was created to combine the results from a between–within analysis into a more conceptually meaningful solution.

As in contextual analysis, within the APIM, both the individual and the group affect an individual's outcome. However, this model conceptualizes the group effect differently from contextual analysis. Recall that in contextual analysis, the group effect is estimated using the group mean of the independent variable to predict an individual's outcome score. In the APIM, the group effect is viewed not as the effect due to the entire group, because the individual is also a member of the group, but rather as the effect due to the other members of the group. So the group is viewed as the mean of the other members of the group. Mathematically, the equation is $M'_X = (nM_X - X_i)/(n - 1)$, where X_i is the score of person i. Within this model, the effect of X on Y is referred to as an *actor effect*, and the effect of M'_X on Y is referred to as a *partner effect*. In our example, the actor effect estimates the degree to which a person's own number of acts affects his or her level of satisfaction. The partner effect estimates the degree to which the average number of acts by the other group members affects the individual's level of satisfaction.[4]

We feel that M'_X, better captures the influence of the group on the individual than does M_X. Actually, M'_X refers not to the group as a whole but to the others in the group. Although there is some precedent in such a conceptualization (Glomb et al., 1997), it is relatively unique in group research.

The statistical analysis for the APIM proceeds in two steps. First, the between–within analysis is done. The between-groups analysis results in a regression coefficient, b_B, that estimates the effect of the average independent variable on the average outcome variable. The within-group analysis yields a regression coefficient, b_W, that is the effect that the individual's deviation from the group average on the independent variable has on the individual's deviation from the group average on the outcome measure. The actor effect can be estimated using the following formula:

$$a = \frac{b_B + (n - 1)b_W}{n},$$

and the partner effect by

$$p = \frac{(n - 1)(b_B - b_W)}{n}.$$

Again, the actor effect estimates the effect that a person's own score on the independent variable has on that person's outcome measure, and the partner effect estimates the effect that the other group members' scores on the independent variable have on the person's outcome measure.

If group sizes are equal, the coefficients of the model can be transformed into the contextual analysis in the following way: The effect of the X is $(a - p)/(n - 1)$, and the effect of the group mean, or M_X, is $pn/(n - 1)$, where a and p are actor and partner effects, respectively. Note that the test of M_X in contextual analysis is the same as the test of M'_X. From the point of view of the APIM, in the contextual analysis, the effect of the person or X is underestimated because part of the person effect is contained in the mean, so the two are, to some extent, collinear.

Significance tests for the actor and partner effects are based on a combination of the standard errors of b_B and b_W, designated as s_B and s_W. The standard error of a is the square root of

$$\frac{s_B^2 + (n - 1)^2 s_W^2}{n^2},$$

and the standard error of p is the square root of

$$\frac{(n - 1)^2 (s_B^2 + s_W^2)}{n^2}.$$

The Satterthwaithe (1946) approximation is used to estimate the degrees of freedom. The degrees of freedom for actor are

$$\frac{[s_B^2 + (n - 1)^2 s_W^2]^2}{\dfrac{s_B^4}{df_B} + \dfrac{(n - 1)^4 s_W^4}{df_W}},$$

and for partner they are

$$\frac{(s_B^2 + s_W^2)^2}{\dfrac{s_B^4}{df_B} + \dfrac{s_W^4}{df_W}}.$$

This approximation generally results in fractional degrees of freedom.

For the example, the estimated actor effect is [0.126 + 7(1.025)]/8, which equals 0.913, and the estimated partner effect is 7[0.126 − 1.025]/8, which equals −0.787. So a person who had more acts was more satisfied, but a person whose fellow group members had more acts was less satisfied. The estimated standard error for actor is 0.105, and the estimated standard error for partner is 0.226. The degrees of freedom are 55.8 for actor and 46.5 for partner and are indeed fractional. The resulting significance test results are $t(55.8) = 4.69$, $p < .001$, for actor and $t(46.5) = -3.48$, $p = .001$, for partner.

Despite several important advantages of the APIM over previous models, there are some limitations. First, the analysis is not straightforward because it requires one to compute the between-groups and within-group regressions and then combine the results from those analyses. Second, if one wanted to estimate a restricted model—for instance, one with actor effects for some independent variables and partner effects for others—it would be difficult to do so. Third, the formulas above are based on the assumption that group sizes are equal. If that were not the case, it would not be clear how to generalize the formulas to the unequal group size case. It is for these reasons that we turn to multilevel modeling as a more general alternative.

[4] The APIM clarifies the differences between dyadic and group research. For dyadic research, the "other group members" is just one person, so the "other" has a specific referent. For groups of three or more, the "other" is a statistical average of scores.

Traditional Multilevel Analysis

Multilevel models, also called hierarchical linear models, random coefficient models, and mixed models, provide a relatively new data-analytic approach that can be used when data have a hierarchically nested structure. Data from small-group research are hierarchically structured when each individual is a member of one group. As we have emphasized throughout this article, there are two possible levels of analysis in group data: individual and group. In this section, we describe what might be termed the *traditional multilevel modeling approach for group data* (Bryk & Raudenbush, 1992; Snijders & Bosker, 1999).

In multilevel modeling, there are two stages of estimation (Kenny et al., 1998), the first being at the individual level within group, and the second being at the group level. The first stage of the estimation involves computing an analysis across individuals within each group separately. Thus, in the first stage, for each group, an individual's X (number of acts) is used to predict that individual's Y (satisfaction):

$$Y_{ij} = b_{0j} + b_{1j}X_{1j} + e_{ij}$$

for individual i in group j. Each of these first-stage analyses results in an intercept, or b_{0j} (if the number of acts is centered, the intercept is the average satisfaction score for the group), and a slope, or b_{1j} (the degree to which satisfaction changes as a function of acts within the group). In the example, there were eight groups, so this first-stage analysis is repeated eight times, once for each group.

The second stage of estimation involves aggregating the first-stage results across groups. The simplest second-stage analysis is to average across groups of the first-stage coefficients (both the intercept and the slopes) and then test whether these averages differ from zero. More commonly, in the second stage of estimation, the coefficients that were estimated at the first stage (b_0 and b_1) are each treated as outcome variables, and any variable that varies only at the group level is used to predict those coefficients. Say that in the example, some groups were all male, and others were all female. The second-stage analysis would test whether male or female groups were more satisfied on average (gender predicting the first-stage intercepts) and whether the relationship between acts and personal satisfaction was stronger in male versus female groups (gender predicting the first-stage slopes). It is possible, although not always done, to treat the mean of the X variable for each group, or M_X, as a predictor in this second-stage analysis (Bryk & Raudenbush, 1992).[5]

In the most general multilevel model, the coefficients from the first-stage analyses (intercepts and slopes) are allowed to vary across groups. However, constraints can be placed on the model. By definition, in small-group research, group size is small, so there may not be enough individuals within each group to estimate the slopes and intercept separately for each group. For instance, if the group size were four and there were four individual-level predictors, the regressions could not be estimated for each group. This situation is typical in small-group research, so it is advisable not to allow the slopes to vary across groups but to allow the intercepts to vary. It is this variation in the intercepts (akin to group means) that models the nonindependence in the groups. Assuming that the predictor variables, or Xs, are centered about their grand means, one can compute the ratio of the variance of the intercepts to the sum of variance of the intercepts and the error variance. This ratio represents the proportion of variance due to groups and is similar to the intraclass correlation (more technically, a partial intraclass correlation for which the effects of the predictor variable are partialed).

A key distinction in multilevel models is fixed versus random effects. The fixed effects in the model are the average intercept and the average slope for each predictor variable from the second stage of estimation. To say that a term is *random* implies that it varies and therefore is associated with a variance. In principle, there is the variance of the intercepts, the variance of the slopes for each X variable, and the error variance. Although it is possible to treat the slopes for X from each group as a random variable (i.e., the effect of X is different for each group), with small-group data there are usually not enough observations to do so. Thus, the two random effects in small-group research are the intercept and the error.

Researchers sometimes center each individual's predictor score about that individual's group's mean (rather than centering about the grand mean). In this case, the first-step predictor score is $X - M_X$ and not X. This measure is analogous to the within-group variable of the between–within analysis. If this approach is combined with treating the group mean of X as a Level 2 predictor (see Bryk & Raudenbush, 1992), the resulting analysis is equivalent to the between–within analysis, which, as we described earlier, can be difficult to interpret.

Multilevel models are generally analyzed using either maximum likelihood (ML) estimation or restricted maximum likelihood (REML) estimation. ML estimation uses an iterative solution to derive estimates of both fixed effects (e.g., intercepts and slopes) and random effects (e.g., variances and covariances). REML estimation also uses ML techniques to estimate random effects, but it uses generalized least squares to estimate fixed effects. REML is the default estimation technique in the data analysis program HLM5 (Raudenbush, Bryk, Cheong, & Congdon, 2001) and in PROC MIXED in SAS (Littell, Milliken, Stroup, & Wolfingen, 1996). REML is generally preferred over ML because estimates of fixed effects using ML tend to be biased, particularly with small data sets. In addition, in the case of equal *n*s (i.e., each group has the same number of individuals), the REML solution is equivalent to the standard least squares analysis.

Generally speaking, ML estimation produces a statistic, called the *deviance*, that is a measure of the degree to which the model fits the data. The deviance provides a badness of fit index: The greater the deviance is, the worse the fit is. The deviance measure can be used to test the relative fit of alternative models. Consider first two models, which are the same in their fixed effects but differ in random effects components. In one model, both the intercept and the slope coefficients are treated as random effects, and their variances are estimated. In a simpler model, the slope coefficients are treated as fixed effects, and only the intercept is treated as a random effect. The deviance of the more complex model can be subtracted from that derived from the simpler model, and under the assumption of multivariate normality and the null hypothesis that the constraints of the simpler model are true, this difference has a

[5] An analysis in which both the individual score and the group mean are predictors is the multilevel version of contextual analysis.

chi-square distribution with degrees of freedom being the number of constraints (1 in the example).

The deviance test can be applied to fixed effects in the model only if ML is used to estimate the fixed effects as well as random effects. An example of this situation occurs when a researcher seeks to test whether exclusion of interactions significantly worsens the model fit. Note that both ML and REML approaches provide *t* tests of individual fixed effects. It is only in the case when a test of a subset (greater than 1) of predictors is desired that the deviance test for fixed effects is of interest.

A Revised Multilevel Strategy

Although multilevel modeling has been recommended for the analysis of small-group data, it has rarely been applied (Hoyle et al., 2001). However, we believe that two important changes to the traditional multilevel model should be made for groups. First, the traditional method treats the nonindependence created by groups as a variance (i.e., the variance of the intercepts), and because a variance can never be less than zero, the traditional multilevel approach does not allow for negative nonindependence.[6] However, as we have discussed, the nonindependence that is created by groups sometimes results in a negative correlation between group members: The higher one member scores, the lower the others score. So the multilevel model should allow for the possibility of a negative correlation between group members.

Second, in the traditional multilevel approach (Bryk & Raudenbush, 1992), the researcher models the group effect by using the group mean as a predictor in the second stage of the analysis. We believe that the APIM (Kashy & Kenny, 2000) provides a more appropriate way of modeling the group effect. In that model, as described in this article earlier, a person is affected both by his or her own standing on the predictor variables and by the average of all other members excluding that person on those predictor variables. Thus, the group effect is modeled as a partner effect where the partner effect is the effect that the average standing of other group members on the predictor variable has on the individual's outcome measure. In the example, according to the APIM, a person's level of satisfaction depends on the number of acts (an actor effect) as well as on the average number of acts by the other group members (a partner effect).

To incorporate the APIM into multilevel modeling, rather than using the group mean, M_X, as a second-stage predictor, one uses the group mean excluding the individual, M'_X, as a predictor in the first-stage or individual-level analyses. That is, when one estimates the APIM using multilevel modeling, both a person's own score, X, and the average of the other group members' scores, M'_X, are predictors of the individual's outcome in the first-stage analysis.

Allowing for negative correlations within traditional multilevel computer programs (e.g., HLM) is somewhat awkward, but it can be accomplished (Snijders & Kenny, 1999). As an example, assume that each group is composed of three individuals. In this case, three dummy variables would be created: Person 1 in the group would have a 1 on Z_1 and a zero on the other two variables, Person 2 in the group would have a 1 on Z_2 and a zero on the other two variables, and Person 3 in the group would have a 1 on Z_3 and a zero on the other two variables. We refer to these as *person-indicator variables*, and they would be included as predictors in the first-stage analyses. The model would have no overall intercept, and the fixed effects of Z_1, Z_2, and Z_3 would be set equal and would take on the role of the intercept. There would be no error variance, but the variance of the effects of Z_1, Z_2, and Z_3 would be free and set equal to each other, and these variances would take on the role of error variance. Finally, these *Z*s are intercorrelated, the correlations are set equal, and they would estimate the nonindependence in the data. Because nonindependence is modeled as a covariance and not a variance, it can be negative. For more detail about this approach, the reader might wish to consult Snijders and Kenny (1999).

A simpler strategy that accommodates negative nonindependence and models the group effect as the APIM partner effect is to use PROC MIXED within SAS (Singer, 1998). The statements for SAS are

```
PROC MIXED;
CLASS GROUP;
MODEL Y = X MXPRIME/SOLUTION DDFM = SATTERTH;
REPEATED/TYPE = CS SUBJECT = GROUP;
```

where GROUP is a variable that identifies group membership, Y is an individual's score on the outcome, X is the individual's score on the predictor, and MXPRIME is the mean of X of the other group members. The SOLUTION option in the MODEL statement requests that the estimates for the fixed effects parameters be printed. This gives the estimates of the intercept and the slopes for both the actor effect (the effect of X) and the partner effect (the effect of MXPRIME). The DDFM = SATTERTH option requests the Satterthwaite (1946) approximation to determine the degrees of freedom for the fixed effects (Kashy & Kenny, 2000). In fact, the estimates and tests of this model using restricted ML estimates are the same as the APIM. However, the multilevel model is much more flexible. We note that for complicated large-sample studies, the Satterthwaite approximation can be computationally intensive and very time consuming.

The REPEATED statement treats the individual scores as repeated measures in the group, and CS implies what is called *compound symmetry*, which means that the degree of nonindependence between each pair of group members is equal. Nonindependence is estimated as a covariance and not as a variance.

For the example, the effect of number of acts is $B = 0.913$, and the effect of the average number of acts by others in the group is $B = -0.787$, both of which are identical to the effects estimated by the APIM. The significance test results are also identical. The covariance of errors is -0.140, and the intraclass correlation is estimated by the covariance divided by the variance ($-0.140/1.321 = -.106$), which is statistically significant, $\chi^2(1, N = 64) = 6.05$, $p = .0139$.

Extensions and Complications

In the final section of the article, we extend the model and the statistical analysis in three ways that may be of interest to small-group researchers and theoreticians. We first allow for an interaction between actor and partner effects. Second, we allow for the

[6] As Kenny and Cook (1999) described, some multilevel modeling programs cannot be run when the nonindependence is negative. Others, in essence, treat negative nonindependence as if there was independence, so person is mistakenly treated as the unit of analysis.

effects of other members on the individual (the partner effects) to vary as a function of group size. Third, we allow for the members within a group to have differential effects on the outcomes of the other group members. All three of these modifications allow us to test several interesting hypotheses in the study of groups. Some of the suggestions that we make are preliminary, though detailed, and as far as we know, they have not been previously used in the small-groups literature. Thus, we are attempting to show the kinds of possible analyses that might be done. It is likely that even more powerful and theoretically relevant analyses will eventually be developed, something that we wholeheartedly encourage.

Actor–Partner and Other Interactions

In the basic model that we have proposed, there are two possible effects that one variable might have on another: the effect of X on Y, and the effect of the mean of the other group members' Xs, or M'_X, on Y. The former effect has been called the actor effect, and the latter has been called the partner effect. These two effects may interact in many different ways, and that interaction may be the central focus of a small-group study.

The traditional method for assessing an interaction between effects is to form a product variable (e.g., X times M'_X) and include it as a predictor of the outcome score along with the two main effects (e.g., X and M'_X). Very often, other formulations of the interaction should be considered, depending on the type of interaction expected. One possible actor–partner interaction is that of similarity. For instance, in groups with highly similar members, X and M'_X are very close in value, whereas in groups with dissimilar members, X and M'_X are relatively far apart. An example of a similarity interaction is that a person who is similar to other group members feels more committed to the group than does a person who is dissimilar. If a similarity effect were predicted, then the interaction would be measured using the absolute difference between X and M'_X as a predictor in the analysis. That is, there would be three predictors of the outcome score: X, M'_X, and the absolute difference between the two of them. Note that similarity is itself a mixed variable as long as the group size is three or more. Thus, some individuals in the group are less similar to other group members than are other individuals.

As discussed in Kenny and Cook (1999), there are other ways to form the actor–partner interaction. One possibility is to determine the maximum value of X of the other group members. So, for instance, a person might enjoy being in a group less as function of the most neurotic person in the group: Only one bad apple might make the experience more negative. Alternatively, the appropriate calculation may be the lowest score of the group. For conjunctive tasks (i.e., tasks for which the group outcome depends on the performance of worst member of the group), member satisfaction might be determined by the member who is exerting the least effort.

Partner effect interactions might be particularly useful in determining the extent to which the group is a real group (Kenny & Cook, 1999). Consider a group productivity study in which, in addition to a predictor variable that measures motivation to be productive, there is a measure of each person's identification with the group. Such a study allows for the determination of whether the partner effect for motivation (does having other group members who are highly motivated affect an individual's productivity?) interacts with the individual's level of identification. Presumably, the more that a person identifies with the group, the stronger the influence of the other group members is on that individual. Moreover, a classic small-group hypothesis is that in a more cohesive group influence effects are stronger, so group cohesiveness should increase the size of the partner effects.

Sometimes a reasonable and parsimonious hypothesis is that the actor effect plus the partner effect equals the same value across some interacting variable. For some values of a moderator, the actor effect is larger than the partner effect, and in other cases, the partner effect is larger than the actor effect is. So, for instance, one might think that for new group members, the actor effect is weaker than the partner effect, whereas for old timers, the actor effect is larger than the partner effect.

We seek to constrain the sum of the actor and partner effects to be the same for all group members. This can be accomplished as follows: We denote the basic model by

$$Y = aX + bM'_X + cQX + dQM'_X + fQ + U,$$

where Q is the moderating variable (e.g., length of membership) that influences the relative size of actor and partner effects. If the sum of actor and partner effects is always the same, then it follows that $c + d = 0$. One can impose such a constraint by regressing Y on X, M'_X, $X - M'_X$, and Q, resulting in the following equation:

$$Y = aX + bM'_X + cQ(X - M'_X) + fQ + U.$$

Within this model, the actor effect for a given value of Q equals $a + cQ$, and the partner effect equals $b - cQ$. Thus, the sum of actor and partner effects always equals $a + c$. To evaluate the assumption that the actor plus partner effects sum to the same value, estimate the above equation with either X or M'_X and drop the other variable. If the extra effect were needed in the equation, the assumption would not hold.

The Effect of Group Size

The analysis that we have proposed has assumed that group sizes are equal. This need not be the case. There are two major reasons why group size may vary. First, the research design may involve equal group sizes, but, because of missing data, the group sizes become unequal. In this instance, all groups contain the same number of individuals, but not all the individuals provide data. Alternatively, group sizes may really vary. Thus, unequal group sizes are part of the plan of the study. We first discuss the case in which there are no missing data and group sizes vary because the groups really differ in size. Later in this section, we turn our attention to the missing data question.

If group sizes vary, then the effect of the other group members on the individual can be conceptualized in two ways: It may be that the mean of the other group members affects the individual or that the sum of the other group members affects the individual. The group mean specification implies that the person is affected by the average member, regardless of group size, and the sum specification implies that the person is affected by all members of the group. One can conduct a test of whether group size is important by estimating both models using ML and comparing their deviances, a measure of fit of a multilevel model. The model with the lower deviance provides a better fit to the data.

We can view the mean of other members' scores and the sum of those scores as ends of a continuum, which would be defined by the divisor of the sum. That divisor is $n - 1$ for the mean and 1 for the sum. The best model might lie between these two alternatives. To find the best model, the following variable might be created:

$$\frac{(n_i - 1)(M'_X - M)}{(n_i - 1)^q},$$

where n_i is the size of group *i*, *M* is the mean of *X* across all groups, and *q* is a fixed constant[7] between 0 and 1. When *q* is 1, the formula becomes the mean, and when *q* is 0, the formula becomes the sum. Social impact theory (Latané, 1981) predicts that *q* is a value between 0 and 1. Thus, the value of *q* might be empirically chosen to be the value that has the lowest deviance using ML estimation.

Alternatively, we can just estimate effects for different group sizes and examine the pattern of results. We can also determine whether the degree of nonindependence varies by group size. From such an empirical analysis, we might develop a new theory of the relationship between group size and influence.

The original plan of a study might be one of equal group sizes, but because of missing data, the groups may contain different numbers of individuals. That is, all the members of each group may not be measured. We assume that the data are missing at random (MAR). Normally, MAR does not create bias in sample estimates, but it does for the APIM. Note that M'_X refers to the mean of all scores not including the participant. So if there were missing data on the outcome, there would likely be missing data for *X*, so M'_X cannot be exactly estimated. Because the sample value of M'_X does not exactly equal the population value of M'_X, there is bias in the estimation of the actor and partner effects. We hope that methods to handle such missing data effects can be developed.

Differential Weighting of Group Members

We consider another feature that can complicate the analysis but represents an important social–psychological and personality consideration: Not all members of groups are created equally. Some members have more influence than others, a fact that has been incorporated in computer simulations of group processes (Nowak, Szamrej, & Latané, 1990; Tanford & Penrod, 1984). The analysis that we have proposed so far has assumed that all members are equally influential.

To allow for the fact that some members are more or less influential, we need a measure of influence. Suppose that influence is thought to be a function of status, such that high-status group members are more influential than are low-status members. To model the effect of influence, when computing M'_X we would compute a weighted average, where the weights are determined by status. Because a weight of zero needs to be a meaningful value, status would have to be measured using the ratio level of measurement of status. So instead of computing M'_X by just averaging the scores of the group's other members, we would compute $\Sigma(wX)/\Sigma w$, summing over the other group members (i.e., omitting the individual's own score), where *w* refers to a weight (e.g., status). To determine whether such a differential weighting is necessary, we could have both M'_X and $\Sigma(wX)/\Sigma w$ in the equation and determine whether the latter is statistically significant. That would suggest that differential weighting is advantageous. Alternatively, we could do the analysis twice, once using the simple average, and a second time using the weighted average, and compute the deviance of each model. The model with the lower deviance value would be preferred.

Alternatively, we might measure how influential different types of group members are empirically. Consider a study of four-person groups composed of two men and two women. The value of M'_X could be computed separately for male and female partners, and both values could be included as variables in the analysis. There would then be two partner effects, one for women and one for men. In that way, we could determine whether men had more influence than did women. In much the same way, we can test whether leaders have more influence than followers or whether solo members (one woman in a five-person group) have more influence.[8]

We have considered only differential effects of influence on the part of other group members. It is also probable that there are individual differences in how easily influenced some people are. For instance, there is a literature on gender differences in influenceability (Eagly & Carli, 1981). To model this type of effect, we would need a variable that measures how easily influenced the actor is. We could then allow this variable to interact with the partner effect. In this way, we could see whether some persons were more (or less) easily influenced.

These moderator variables may interact not only with the partner effect but also with the actor effect. For instance, people with greater status may have stronger actor effects. Therefore, the researcher should consider the moderation of actor as well as partner effects. Thus, the complete model is quite complex in that it allows for differential influence and influenceability for both actor and partner effects.

Conclusion

We have discussed many of the special features of small groups that should be taken into account by data-analytic models. Data from groups are often not independent, and some allowance must be made for that nonindependence—not only because it may result in biased statistical tests but also because it may represent unique components of the social interplay that occurs within small groups. A complicating factor that is sometimes overlooked is the possibility that nonindependence can be negative, not as an artifact but because certain group processes lead to differentiation rather than homogenization. Although we have described several different

[7] The perceptive reader may have noticed that the exponent is in the denominator, whereas in social impact theory (Latané, 1981), it is in the numerator. The difference is that social impact theory uses the sum, not the mean, and when we turn the mean into a sum by multiplying through by $n_i - 1$, we have $(n_i - 1)/(n_i - 1)^q$, which equals $(n_i - 1)^{1-q}$.

[8] The above strategy is feasible only if each group has at least two members of each type. But there would be problems if we were to investigate the effect of a variable that did not conform to this requirement. For gender, if some groups were all male or all female or some had only one male or female, we would be unable to use those groups. We believe that there is a solution to this problem that involves the use of dummy variables for each person in the group. However, the exact details of such a solution are as yet unclear.

approaches that can be used to analyze data in which predictor variables vary both within and between groups, we recommend the APIM, which simultaneously accounts for nonindependence, allows for negative nonindependence, and provides the most relevant model for social psychologists who are interested in how individuals influence one another. Multilevel modeling provides an excellent means for estimating the APIM and for testing extensions of this model.

Despite the comprehensiveness of our approach to group data, there are several limitations. First, we have assumed that the outcome variable is measured at the interval level of measurement. This limitation is less problematic than it might appear, as increasingly multilevel models allow for categorical outcome variables. Another serious limitation in this article is that it fails to examine dynamic factors of group interaction. A complete examination of small groups requires an understanding of change (McGrath, Arrow, & Berdahl, 2000). For a discussion of the methods for the sequential analysis of group data, the reader might consult Bakeman and Gottman (1997). Additionally, often the data from groups refer to dyads, not individuals. For example, the outcome variable might be a measure of how often one person in the group talks to another person—therefore, the score refers to a dyad, not an individual. It should be possible to combine an analysis of dyadic outcome scores such as a social relations analysis (Kenny, 1994; Snijders & Kenny, 1999) with the methods in this article. Finally, the network structure of groups is not considered (Wassermann & Faust, 1994). It might be possible to use results from a network analysis (i.e., centrality) to determine influence strength in the group, and those results would feed into the analyses that we have suggested.

Notwithstanding these important limitations, the approach that we have presented offers some unique opportunities for the study of groups. We have combined state-of-the-art statistical methods with current social–psychological theories of groups to optimize the understanding researchers can derive from small-group research. We feel that if researchers were to use this approach, much more might be learned about the processes that generate the data. At the same time, the methods would be revised to take into account the new social–psychological processes discovered by these methods. We hope that these new methods will lead to a strengthening of interest in small groups within psychology as we begin this new millennium.

References

Anderson, L. R., & Ager, J. W. (1978). Analysis of variance in small group research. *Personality and Social Psychology Bulletin, 4,* 341–345.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

Bales, R. F. (1950). *Interaction process analysis.* Reading, MA: Addison-Wesley.

Boyd, L., & Iversen, G. (1979). *Contextual analysis: Concepts and statistical techniques.* Belmont, CA: Wadsworth.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Campbell, L., Simpson, J. A., Kashy, D. A., & Rholes, W. S. (2001). Attachment orientations, dependence, and behavior in a stressful situation: An application of the actor-partner interdependence model. *Journal of Social and Personal Relationships, 18,* 849–871.

Crits-Christoph, P., & Mintz, J. (1990). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology, 59,* 20–26.

Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin, 90,* 1–20.

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

Fletcher, G. J. O., & Fitness, J. (1990). Occurrent social cognition in close relationship interaction: The role of proximal and distal variables. *Journal of Personality and Social Psychology, 59,* 464–474.

Glick, W. (1980). Problems in cross-level inference. In K. H. Roberts & L. Burstein (Eds.), *New directions for methodology of social and behavioral science* (pp. 17–30). San Francisco: Jossey-Bass.

Glick, W. H., & Roberts, K. H. (1984). Hypothesized interdependence, assumed independence. *Academy of Management Review, 9,* 722–735.

Glomb, T. M., Richman, W. L., Hulin, C. L., Drasgow, F., Schneider, K. T., & Fitzgerald, L. F. (1997). Ambient sexual harassment: An integrated model of antecedents and consequences. *Organizational Behavior and Human Decision Processes, 71,* 309–328.

Griffin, D., & Gonzalez, R. (1995). Correlational analysis of dyad-level data: Models for the exchangeable case. *Psychological Bulletin, 118,* 430–439.

Haney, W. (1980). Units and levels of analysis in large-scale evaluation. *New Directions for Methodology of Social and Behavioral Science, 6,* 1–15.

Hoyle, R. H., Georgesen, J. C., & Webster, J. M. (2001). Analyzing data from individuals in groups: The past, the present, and the future. *Group Dynamics, 5,* 41–47.

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 451–477). New York: Cambridge University Press.

Kashy, D. A., & Snyder, D. K. (1995). Measurement and data analytic issues in couples research. *Psychological Assessment, 7,* 338–348.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford Press.

Kenny, D. A. (1996). Models of nonindependence in dyadic research. *Journal of Social and Personal Relationships, 13,* 279–294.

Kenny, D. A., & Cook, W. (1999). Partner effects in relationship research: Conceptual issues, analytic difficulties, and illustration. *Journal of Personal Relationships, 6,* 433–448.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99,* 422–431.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233–265). Boston: McGraw-Hill.

Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology, 48,* 339–348.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review, 19,* 195–229.

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36,* 343–356.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfingen, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychology Review, 4,* 95–105.

Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Boston: Allyn & Bacon.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion. *Psychological Review, 97,* 362–376.

Pierro, A., & Livi, S. (2000). [IPA and SYMLOG observations of group members]. Unpublished raw data, University of Rome, Sapienza.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2001). *HLM5: Hierarchical linear and nonlinear modeling.* Chicago: Scientific Software International.

Satterthwaite, F. W. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2,* 110–140.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23,* 323–355.

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

Snijders, T. A. B., & Kenny, D. A. (1999). The social relations model for family data: A multilevel approach. *Personal Relationships, 6,* 471–486.

Tanford, S., & Penrod, S. (1984). Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin, 95,* 189–225.

Wassermann, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge, England: Cambridge University Press.

Received September 22, 2000
Revision received November 14, 2000
Accepted November 14, 2000 ■