Using multilevel random coefficient modeling to analyze social interaction diary data

John B. Nezlek

College of William & Mary

ABSTRACT

This article provides a rationale for using multilevel random coefficient modeling (MRCM) to analyze social interaction diary data, and describes how to conduct such analyses. This description includes how to examine relationships between individual differences in social interaction and other trait-level individual differences such as personality scales. Analyzing relationships among different ratings of interactions and analyzing differences among types of interactions are also described, including techniques to examine how relationships between trait-level individual differences and reactions to interaction vary across different types of interactions (e.g., interactions with different relational partners). Different types of analyses are illustrated using data from actual studies. Although this article focuses on the analysis of data collected using variants of the Rochester Interaction Record, the techniques described can be used to analyze data collected using other methods and protocols.

KEY WORDS: multilevel modeling • social interaction diaries • subjects analyses

Interest in day-to-day social interaction has increased considerably over the past two decades, and much of this research has used some form of diary with which people describe the social interactions they have. Unfortunately, the data structures generated in such studies can be quite complex and can present analysts with numerous challenges. Moreover, such challenges may be particularly daunting for analysts whose primary training and experience has been with ordinary least squares (OLS) analyses of data generated using more traditional methods such as surveys or laboratory studies. This

All correspondence concerning this article should be addressed to John B. Nezlek, College of William & Mary, Department of Psychology, PO Box 8795, Williamsburg, VA 23187–8795, USA [e-mail: jbnezl@wm.edu].

Journal of Social and Personal Relationships Copyright © 2003 SAGE Publications (www.sagepublications.com), Vol. 20(4): 437–469. [0265–4075 (200308) 20:4; 034145]

is because OLS techniques, such as analysis of variance and regression, are not the best techniques for analyzing the type of data generated in a social interaction diary study.

This article discusses different ways of analyzing data generated in social interaction diary studies. OLS analyses are described briefly, a rationale for using a newer technique, multilevel random coefficient modeling (MRCM) is provided, and how to use MRCM to analyze the data collected in social interaction diary studies is described. The article is not intended for the statistically sophisticated reader; rather, it is intended for analysts whose primary experience is with OLS regression and ANOVA. Moreover, the article was designed to provide inexperienced analysts with sufficient background information and technical detail to allow them to conduct MRCM analyses of a diary data set. Nevertheless, readers who are somewhat familiar with MRCM may find certain more technically focused sections useful, and such readers may find it useful to read some sections alone.

Although various methods can be used to study social interaction, this article focuses on the analysis of data generated by the use of the Rochester Interaction Record (RIR; Wheeler & Nezlek, 1977), because many studies of naturally occurring social interaction have used variants of the RIR. Nevertheless, the issues discussed in this article are applicable to the analysis of data generated by other methods, and a more general discussion of MRCM analyses of diary data structures data structures can be found in Nezlek (2001). Also, see Gable and Reis (1999) and Snijders and Kenny (1999) for discussions of MRCM analyses of data structures commonly found in research on personal relationships. General introductions to MRCM can be found in Bryk and Raudenbush (1992), Kreft and de Leeuw (1998), and Snijders and Bosker (1999).

Participants in RIR studies use a standardized form to describe the social interactions they have each day, usually for a 1–2-week period. These descriptions typically include the date, time of onset, and duration of the interaction, the sex, initials, and role relationships of the other people present, and various ratings and descriptions of the event. Analyses of RIR data typically concern three types of questions.

1. Perhaps the most commonly investigated question concerns relationships between characteristics of social interaction and individual differences such as the relationship between depression and quality of interaction (e.g., Nezlek, Hampton, & Shean, 2000; Zuroff, Stotland, Sweetman, Craig, & Koestner, 1995).
2. A second type of question concerns within-person relationships such as differences between interactions with friends and lovers (Nezlek, 1995) or differences between interactions that involve lies and those that do not (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996).
3. A third type of question concerns how a within-person relationship such as relative reliance on close versus distant friends varies as a function of an individual difference variable such as depression (Nezlek, Imbrie, & Shean, 1994) or sex (Wheeler & Nezlek, 1977).

Such relationships can be examined in terms of two types of measures. The first type of measure, and the one that has received the most attention, consists of ratings of interactions such as enjoyment and intimacy. The second type of measure consists of measures of interaction quantity and the distribution of interactions such as interactions per day and the percent of interaction involving romantic partners. This article describes the type of analyses appropriate for examining the three types of questions mentioned earlier for ratings of interaction (referred to collectively as quality) and for interaction quantity.

Regardless of the analytic strategy one uses, it is critical to take into account the possibility (if not the likelihood) that the relationships between constructs at one level of analysis may not be the same as the relationships between these same constructs at another level of analysis. For example, assume a researcher is interested in the relationship between the intimacy people find in their interactions and how satisfying their interactions are – the two scales used by Wheeler and Nezlek (1977) in the study that introduced the RIR. One way to study this relationship is represented by a question such as: Do people who experience more intimacy in their interactions find their interactions to be more satisfying? A second way would be represented by the question: Are interactions that are more intimate more satisfying?

The first question involves a between-person analysis, and the second involves a within-person analysis. Mathematically speaking, these levels of analysis are independent: Any sort of relationship that can exist at one level can co-exist with any sort of relationship at the other. This independence is illustrated by the two sets of hypothetical data presented in Table 1. Assume that three people have each rated how satisfying and intimate they found five interactions to be. In the first set, there was a positive relationship between these two constructs at the between-person level, a relationship represented by the numbers in the bottom row labeled 'Mean'. People who had more intimate interactions tended to have more satisfying interactions. In contrast, for all three people, there was a negative relationship between intimacy and satisfaction at the interaction level. For any particular interaction, greater intimacy was associated with diminished satisfaction. In the second data set, the between-person relationship was negative, whereas all the within-person relationships were positive. It is relatively easy to recognize from these examples that any combination of relationships can exist simultaneously at both levels of analysis.

Such combinations of relationships naturally beg the question of which level of analysis is correct. The answer is that neither is correct de facto. Questions about interaction-level phenomena are answered by analyses at the interaction (i.e., within-person) level, whereas questions about person-level phenomena are answered by analyses at the between-person level. Analysts simply need to be certain that the levels of analysis implicit in their questions correspond to the levels of analysis represented in their analyses.

TABLE 1
Relationships between satisfaction and intimacy at different levels of analysis

	Person 1		Person 2		Person 3	
Interaction	**Sat.**	**Int.**	**Sat.**	**Int.**	**Sat.**	**Int.**
Positive Between-Person Relationship						
Negative Within-Person Relationship						
1	1	7	2	8	3	9
2	2	6	3	7	4	8
3	3	5	4	6	5	7
4	4	4	5	5	6	6
5	5	3	6	4	7	5
Mean	3	5	4	6	5	7
Negative Between-Person Relationship						
Positive Within-Person Relationship						
1	4	3	5	2	6	1
2	5	4	6	3	7	2
3	6	5	7	4	8	3
4	7	6	8	5	9	4
5	8	7	9	6	10	5
Mean	6	5	7	4	8	3

Types of analyses

Ordinary least squares analyses

It is difficult to use traditional OLS techniques to analyze the data generated in most social interaction diary studies for various reasons. First, unless they are arbitrarily constrained, participants in RIR studies invariably have different numbers of social interactions. Such data cannot be analyzed using traditional OLS-based repeated measures analyses of variance in which interactions are treated as a repeated measure, because OLS ANOVA requires that all units of analysis have equal numbers of observations. Moreover, the validity of within-subject regression analyses may be compromised by the fact that the observations (i.e., interactions) being analyzed are not independent.

The most widely used solution to the problems created by unequal numbers of observations has been the use of aggregation analyses. In these analyses, individual-level summary measures are calculated and are then analyzed using standard OLS techniques. For example, average intimacy in interaction or number of interactions per day are calculated, for all interactions or for interactions of a certain type such as interactions with romantic partners or close friends. Such summary measures can be analyzed with ANOVA or regression just like any individual difference measure. Such analyses were introduced by Wheeler and Nezlek (1977) and were described in detail in Nezlek and Wheeler (1984). Although the bulk of social interaction diary research has relied on within-person means, other types of within-person aggregates have been used. For example, Hodgins, Koestner, and Duncan (1996) calculated within-subject correlations among different ratings of interactions and then correlated these correlations with other individual difference variables.

Although aggregation analyses solve the problems associated with unequal numbers of observations, they create other problems. For example, other considerations being equal, summary measures based on more observations (i.e., more interactions) tend to be more reliable than measures based on fewer observations. Although this difficulty can be remedied by weighting summary measures by their reliabilities (e.g., Kenny, Kashy, & Bolger, 1998), OLS analyses of aggregated measures (weighted or not) suffer from a fundamental shortcoming in that they treat (model) these summary measures as fixed effects when they should be modeled as random effects.

Whether an effect should be modeled as fixed or random requires consideration of its inference space, the population a coefficient or summary measure is meant to describe. See Littell, Milliken, Stroup, and Wolfinger (1996, pp. 230–234) for a concise discussion of this topic. For example, in most studies of day-to-day social interaction, little if any importance is placed on the specific interactions that are described. Interactions are (presumably) randomly sampled from each person's population of interactions and are meant to represent participants' typical lives. Presumably, summary statistics based on samples of other interactions would be just as valid (although not exactly the same) as those based on the sample collected. This means that such summary statistics are random in that they are sampled from a person's population of possible statistics, and it means that these statistics have error terms. This sampling and the associated error constitute a prima facie case for considering such statistics as random, not fixed, effects.

Unfortunately, no matter what weighting system is used, OLS techniques cannot simultaneously model the random error associated with sampling individuals and the random error associated with sampling interactions. Moreover, the impact of this inability on significance tests is unpredictable – sometimes modeling random effects as fixed will make tests of fixed effect more conservative and sometimes it will make them more liberal.

Multilevel random coefficient models

Social interaction diary studies produce what are referred to as multilevel or nested data structures. These terms refer to the fact that observations at one level of analysis (e.g., interactions) are nested within another level (people). Some readers may be familiar with multilevel analyses in which individuals are nested within other units (e.g., groups, classrooms). Such analyses are conceptually similar to the analyses described here; however, in the present analyses, individuals (not groups or classrooms) serve as the organizing units. Moreover, the types of questions commonly addressed in social interaction diary studies can be conceptualized in terms of different levels of analysis and combinations of levels of analyses within a multilevel framework. Current thinking suggests that multilevel random coefficient modeling (MRCM) is the best way to analyze multilevel data structures. The specific approach to MRCM discussed in this article is described by Bryk and Raudenbush (1992), and the software that implements this approach is HLM (Version 5; Raudenbush, Bryk, Cheong, & Congdon, 2000). Although MRCM relies on maximum likelihood procedures to estimate coefficients, in some ways, the analyses are functionally equivalent to calculating regression equations at one level of analysis (e.g., within-person) and then using the coefficients from these equations as dependent measures at the next level of analysis (e.g., between-person).

MRCM provides more accurate parameter estimates than comparable OLS analyses for various reasons. First, MRCM is able to model simultaneously two sources of random error, the error associated with sampling people and the error associated with sampling interactions. In addition, when estimating parameters, HLM uses a Bayesian procedure that includes weighting units of analysis by the reliability of the coefficients for each unit, a process known as precision weighting. This is part of the procedure HLM uses to distinguish true and error variance, which provides more accurate significance tests. For general discussions of the advantages of MRCM over OLS analyses, see Bryk and Raudenbush (1992) and Kreft and de Leeuw (1998).

Although it may be unusual to use equations and models to describe traditional OLS analyses, it is common practice when describing multilevel analyses, and such descriptions are provided in this article. Multilevel analyses are inherently more complex than most OLS analyses because multiple levels of analysis are being considered simultaneously, and it is helpful to make explicit exactly what is being done at each level of analysis. The presentation of multilevel analyses can be simplified somewhat, however, by using the nomenclature that is standard for multilevel analysis, and the models and equations in this article are described using this standard nomenclature. Readers are strongly urged to follow these conventions when describing their own analyses.

Before discussing specific types of analyses, two aspects of MRCM (random error and centering) that are different from traditional OLS analyses are discussed briefly, with more detailed discussion of random error following the discussion of analyses. Regarding random error, in MRCM, two parameters are estimated for each effect, fixed and random effects. Fixed effects concern hypotheses about whether coefficients are different from 0 and will probably be the primary interest for the overwhelming majority of analysts. Such coefficients can represent a wide variety of constructs: means, difference scores, covariances, interaction effects, and so forth. Part of the power of MRCM lies in its flexibility. Within MRCM, variables can be collected, created, and transformed to represent a virtually limitless array of constructs.

In contrast, random effects concern hypotheses about the nature of the variability of an effect. Random error terms can be tested for significance, and if a random term is not significant it should be deleted from a model. There is no need to estimate it. It is *absolutely critical* to note, however, that a nonsignificant random error term *does not mean* that there is no variability in an effect. The significance test for a random error term tests if the random error for an effect can be reliably estimated, and even if random error cannot be reliably estimated, an effect may still vary nonrandomly. Random error is discussed in Nezlek (2001) and later in this article.

Centering refers to the value assigned to an intercept that is used when estimating the other coefficients in a model. For most OLS regression analyses, variables are ‘mean-centered.’ Coefficients reflect deviations from sample means, and the intercept represents the sample mean and is also the estimated score for an observation with a mean score on all the measures in a model. For the many analysts who are primarily interested in standardized coefficients, the intercept is typically of little importance, and standardized models have no intercepts (or more accurately, the intercepts are always 0).

The situation in MRCM is not so straightforward. First, in MRCM only unstandardized coefficients are estimated (a topic discussed later). Second, different centering options are available for predictors at different levels of analysis. At the highest level of nesting (the person level for the analyses discussed in this article), predictors can be either grand-mean centered (centered around the mean of the sample) or uncentered. When predictors are grand-mean centered, the intercept represents the score on the dependent measure for a person who is at the mean on the predictor variable(s), and the coefficient for a predictor estimates how a dependent variable deviates from the mean as a function of the predictor. This is the same as in unstandardized OLS models. When predictors are uncentered, the intercept represents the score on the dependent measure for a person who has a 0 on the predictor variable(s), and the coefficient estimates how a dependent variable deviates from 0 as a function of the predictor.

At lower levels of nesting (the interaction level for present purposes, or level 1), predictors can be either uncentered, grand-mean centered, or group-mean centered. When level 1 predictors are uncentered, the intercept represents the predicted score for a dependent measure when a predictor is 0. In addition, the coefficient for a predictor (called a slope) that is estimated for each person is centered around 0. That is, the slope represents how a predictor varies from 0 for a particular person. When predictors are grand-mean centered, the intercept represents the predicted score for a dependent measure when a predictor is at the grand-mean (the sample mean), and the slope represents how a predictor varies from the grand mean for a particular person. When predictors are group-mean centered, the intercept represents the predicted score for a dependent measure when a predictor is at the mean for each person, and the slope represents how a predictor varies from each person’s mean on that predictor.

Regardless of the type of analysis, the key to interpreting the results of HLM analyses is to generate predicted (or estimated) values. Analysts are strongly encouraged to ‘plug in’ different values in both interaction-level

(level 1) and person-level (level 2) models. This is done by multiplying different values for variables by the coefficients estimated for these variables. This will be done for each of the examples presented below.

Within the terminology of multilevel modeling, this article focuses on two-level models because two levels should be sufficient for most applications. Three-level models are discussed, however, in terms of special applications after the basic principles of multilevel modeling are presented using the two-level models. Although analyses of ratings of interactions (quality) and interaction quantity have much in common, there are sufficient differences between the two to warrant presenting them separately.

Interaction quality

Analyses of data reported by Nezlek et al. (2000) are used to illustrate some analysis of interaction quality. In this study, 43 participants (19 of whom were clinically depressed) maintained a variant of the RIR for two weeks. They rated each interaction in terms of how enjoyable and intimate it was and how much influence they felt they had over the interaction. For present purposes, enjoyment is the prime dependent measure.

Basic models

For such analyses, interactions are treated as nested within people, and the basic level 1 (or interaction level or within-person) model is:

$$y_{ij} = \beta_{0j} + r_{ij}.$$

In such a model, y_{ij} is a rating for each interaction (subscripted i) for each participant (subscripted j), β_{0j} is a random coefficient (an intercept) representing the mean of y across all interactions, r_{ij} represents error, and the variance of r_{ij} represents the level 1 residual (error) variance. In essence, this model estimates within-person means and an error term for a rating.

In multilevel modeling, the coefficients from one level of analysis are passed on to the next. For present purposes, this means that individual differences in interaction-level phenomena are analyzed at level 2. The basic level 2 (or person-level) model is:

$$\beta_{0j} = \gamma_{00} + u_{0j}.$$

In this model, γ_{00} represents the grand mean of the person-level means (β_{0j}s) from the interaction-level model, u_{0j} represents the error of β_{0j}, and the variance of u_{0j} constitutes the level 2 residual (error) variance.

Together, two models such as these comprise what is called a ‘totally unconditional’ or ‘null’ model. The phrase totally unconditional refers to the fact that no term other than the intercept is included at each level. Although such models typically do not test hypotheses per se, they describe how much of the total variance of y is at each level. The total variance of y is the sum of the variances at each level. In a two-level model, this is the sum of the variance of r_{ij} and of u_{0j}, and the distribution of the total variance of y suggests the levels at which further analyses might be productive. For example, if all the variance in a measure is at the interaction level, it may be difficult to model between-person differences.

The totally unconditional analysis of enjoyment of the Nezlek at al. data estimated the between-person variance (the variance of u_{0j}, the random error term associated with the intercept, β_{0j}, the mean enjoyment for a person) to be 2.16 and the within-person variance (the variance of r_{ij}, the level 1 random error term) to be 3.32. This distribution of variability suggested that there were meaningful differences both between people and between interactions.

Analyzing individual differences in interaction

Relationships between mean ratings of interactions and individual differences such as depression can be examined by adding terms to the level 2 model. For example, Nezlek et al. examined the relationship between depression and how enjoyable people found their interactions to be. They used a contrast-coded variable (–1 = *depressed*, 1 = *not depressed*) to represent diagnosis in the following level 2 model (with enjoyment as the y variable):

$$\beta_{0j} = \gamma_{00} + \gamma_{10}\,(\text{Diagnosis}) + u_{0j}.$$

The relationship between mean enjoyment (β_{0j}) and diagnosis is tested by the significance of the γ_{10} (Diagnosis) coefficient. To determine if the depression effect on enjoyment varied as a function of sex and cohabitation, terms representing participant sex, cohabitation, and the interactions of sex and cohabitation with depression were included in other models. See Aiken and West (1991) for a discussion of how to test interactions within regression analyses. These interaction analyses are not discussed here. As reported by Nezlek et al., the mean enjoyment (γ_{00}) for the sample was 6.16, and the γ_{10} (Diagnosis) coefficient was .87, which was significantly different from 0. This represented a difference of 1.74 between depressed and nondepressed participants, a difference obtained by comparing the estimated means for depressed (5.29 = 6.16 + [–1 *.87]) and nondepressed participants (7.03 = 6.16 + [1 *.87]).

Analyzing relationships between ratings of interactions

Just as terms can be added to level 2 models to examine person-level relationships, terms can be added to level 1 models to examine interaction-level phenomena. For example, the following model examines the relationship between influence and enjoyment:

$$y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Influence}) + r_{ij}.$$

In such a model, y_{ij} is a rating for each interaction (subscripted i) for each participant (subscripted j), β_{0j} is a random coefficient (an intercept) representing the intercept of y, β_{1j} (Influence) is a random coefficient (referred to as a slope to distinguish it from an intercept) representing the relationship between enjoyment and influence, and r_{ij} represents error. In essence, this model estimates a regression equation for each person (with unstandardized coefficients) describing the relationship between enjoyment and influence.

The hypothesis that the mean slope (i.e., the mean relationship between enjoyment and influence) is 0 is tested at level 2 with the following model:

$$\beta_{1j} = \gamma_{10} + u_{0j}.$$

In this model, γ_{10} represents the mean of the slopes (the β_{1j}s) from the interaction-level model and u_{1j} represents the error of β_{1j}. If γ_{10} is significantly different from 0, then the null hypothesis is rejected. Analyses of the data from Nezlek at al. carried out for this article found that the enjoyment–influence slope was significant (γ_{10} = .27), as was the enjoyment–intimacy slope (γ_{10} = .40). For these analyses, intimacy and influence were group mean centered, which means that each person's coefficient was based on deviations from his or her own mean intimacy or influence.

These coefficients are interpreted like unstandardized regression coefficients. On average, for every 1.0 the perceived influence in interaction was above the mean influence for a person, enjoyment was .27 above the mean, and for every 1.0 influence was below the mean, enjoyment was .27 below that person's mean. For intimacy, the corresponding changes were .40.

Similar to OLS multiple regression, more than one variable can be included in a level 1 model. For example, enjoyment can be predicted by both intimacy and influence:

$$y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Intimacy}) + \beta_{1j}\,(\text{Influence}) + r_{ij}$$

This would lead to the following level 2 model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}.$$

This analysis found that the slope for intimacy was significant and approximately the same size as in the previous analysis ($p < .01$, γ_{10} = .34). The slope for influence also remained significant, although it was smaller than in the previous analysis ($p < .01$, γ_{20} = .14). The differences in the coefficients from these two analyses were due to the fact that intimacy and influence were related ($p < .01$).

In addition to examining the relationships among enjoyment, intimacy, and influence, per se, the strength of the enjoyment–intimacy and enjoyment–influence relationships can be compared using multi-parameter tests for fixed effects (Bryk & Raudenbush, 1992, pp. 48–56). The null hypothesis of such a test is that the two slopes (γ_{10} and γ_{20}) are equal. A comparison of the enjoyment–intimacy and enjoyment–influence slopes from the Nezlek et al. data found that the intimacy slope was greater than the influence slope ($p < .01$). Enjoyment covaried more strongly with intimacy than with influence.

Individual differences in level 1 slopes (e.g., individual differences in the relationships between enjoyment and intimacy and between enjoyment and influence) can also be examined. In the terminology of multilevel modeling, such analyses are sometimes referred to as 'slopes as outcomes' analyses (because the slope from level 1 becomes an outcome or dependent measure at level 2) or as 'cross-level interactions' (because a relationship at one level of analysis, level 1, varies as a function of a variable at another level of analysis, level 2). Such relationships can be examined with a model that is structurally similar to that used to examine individual differences in intercepts, except that the dependent measure is now a slope rather than an intercept. For example, to determine if the slopes from the previous analysis varied as a function of depression, the following model was analyzed:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{Diagnosis}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\,(\text{Diagnosis}) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\,(\text{Diagnosis}) + u_{2j}.$$

The results of these analyses were clear. Diagnosis did not moderate the enjoyment–intimacy slope. The coefficient (γ_{11}) for diagnosis in the intimacy slope equation was not significantly different from 0 ($\gamma_{11} = -.01$). In contrast, diagnosis did moderate the enjoyment–influence slope ($p < .01$, $\gamma_{21} = .15$). In this analysis, the mean slope for influence was .17. The estimated mean slope for depressed people was .02 (.17 + [–1 *.15]), whereas it was .32 (.17 + [1 *.15]) for the nondepressed. This means that for the depressed, every 1.0 increase in influence was accompanied by only a .02 increase in enjoyment, whereas for the nondepressed, every 1.0 increase in influence was accompanied by a .32 increase in enjoyment. Moreover, the depression effects for the two slopes can also be compared using the chi-squared test of fixed effects described earlier. This is a particularly rigorous test compared to examining the significance of each relationship separately, and, for these data, the depression effects for the two slopes were significantly different ($p < .01$).

This example used a dichotomous measure as a moderator; however, continuous measures can be used, and continuous and dichotomous measures can be used together. Examples of such analyses can be found in Nezlek (1999), Nezlek and Derks (2001), Nezlek and Leary (2002), and Nezlek, Richardson, Green, and Schatten-Jones (2002). Finally, it is important to note that individual differences in slopes can be examined even though a mean slope is not significantly different from 0. Even if the mean slope for a sample is 0 there may still be meaningful variability in the sample, for example, if half of the sample has a positive slope and half has a negative slope.

Comparing different types of interactions

The previous example considered all interactions together. It is possible, however, for (1) mean reactions to vary across types of interaction, (2) the relationship between a person-level variable and an interaction rating to be different for different types of interactions, and (3) the relationship between reactions to vary across type of interaction. Although examining such differences relies on techniques similar to those used to examine the interaction-level relationship between two continuous measures (i.e., the previous examples), because type of interaction is represented categorically, there are important differences between the types of analyses.

Generally speaking, interactions can be classified using categories that are mutually exclusive or not exclusive. An example of a mutually exclusive system is the sexual composition of an interaction: same-, opposite-, or mixed-sex. Such a system is mutually exclusive because from the perspective of the participant, a particular interaction involves either same-sex others, opposite-sex others, or members of both sexes – it cannot be two of these simultaneously. An example of a nonexclusive system is one that categorizes interactions on the basis of the presence or absence of relational partners such as romantic partners, friends, co-workers, and so forth. Such a system is not exclusive because a particular interaction can involve more than one type of relational partner (e.g., both a romantic partner and a friend may be present), making it impossible to classify such interactions as only one type or another. Analyzing mutually exclusive categorical systems is more straightforward than analyzing nonexclusive systems and is discussed first.

Mutually exclusive categories

Comparisons of same-, opposite-, and mixed-sex interactions will be used to illustrate multilevel analyses of a mutually exclusive categorical system because distinguishing interactions on this basis has been informative in past research (at least in studies of collegiate samples). Some of these analyses will be illustrated by analyses of data presented in Nezlek and Leary (2002). In this study, 164 participants maintained a variant of the RIR for one week, and they completed a series of questionnaires designed to measure traits related to self-presentational concerns and motives. Intimacy of interaction is the dependent variable in these analyses. Within a regression framework, categorical variables can be represented using either dummy codes (e.g., 0 = *not*, 1 = *yes*) or contrast codes (e.g., –1 = *not*, 1 = *yes*); although these two systems are similar in their focus, there are important differences between analyses relying on the two types of codes. Each provides advantages the other does not.

Dummy codes. When using dummy codes, the basic interaction model is:

$$y_{ij} = \beta_{1j}\,(\text{Same}) + \beta_{2j}\,(\text{Opposite}) + \beta_{3j}\,(\text{Mixed}) + r_{ij}.$$

For this analysis, three variables are assigned to each interaction. *Same* is a dummy-coded variable (0, 1) representing whether or not all the other people present were of the same sex as the diary keeper, *Opposite* is a dummy-coded variable representing whether or not the other people were of the opposite sex, and *Mixed* is a dummy-coded variable representing if the other people present included both men and women. These three variables are included in a model that has no intercept, and the three predictors are entered uncentered (including the intercept or centering would produce linear dependence among the three variables). Such a scheme produces three coefficients that represent the mean intimacy for same-, opposite-, and mixed-sex interactions. This can be verified by generating predicted values for same-sex interactions (multiplying the coefficients by 1, 0, 0 respectively), opposite-sex interactions (0, 1, 0), and mixed-sex interactions (0, 0, 1). Note that these coded variables are assigned to each interaction.

Within such a framework, comparisons among mean ratings for the three types of interactions can be made at level 2 as follows. The basic level 2 (person-level) model generated by this dummy-coded analysis is:

$$\begin{aligned} &\text{Same:} && \beta_{1j} = \gamma_{10} + u_{1j} \\ &\text{Opposite:} && \beta_{2j} = \gamma_{20} + u_{2j} \\ &\text{Mixed:} && \beta_{3j} = \gamma_{30} + u_{3j}. \end{aligned}$$

In this model, γ_{10}, γ_{20}, and γ_{30} represent the mean intimacy (across all participants) for same-, opposite-, and mixed-sex interactions, respectively, and these coefficients can be compared (i.e., the equality of the means can be tested) using the multi-parameter test described previously. For the Nezlek and Leary data, mean intimacy scores for same-, opposite-, and mixed-sex interactions (i.e., γ_{10}, γ_{20}, and γ_{30}) were 3.77, 4.51, and 3.77, respectively. Comparisons of these means found that opposite-sex interactions were more intimate than same- or mixed-sex interactions (*p*s < .01), whereas the intimacy of same- and mixed-sex interactions did not differ significantly.

Differences across type of interaction in the strength of the relationship between a person-level variable and a mean rating of interaction can also be examined. Nezlek and Leary (2002) examined the relationship between ratings of interactions and a factor they labeled *Impression Construction Positivity* (ICP), which represented people's desire to be seen positively by others. Relationships between ICP and the intimacy of same-, opposite-, and mixed-sex interaction were examined using the following model:

$$\begin{aligned} &\text{Same:} && \beta_{1j} = \gamma_{10} + \gamma_{11}\,(\text{ICP}) + u_{1j} \\ &\text{Opposite:} && \beta_{2j} = \gamma_{20} + \gamma_{21}\,(\text{ICP}) + u_{2j} \\ &\text{Mixed:} && \beta_{3j} = \gamma_{30} + \gamma_{31}\,(\text{ICP}) + u_{3j}. \end{aligned}$$

These analyses found that ICP was significantly related only to intimacy of same-sex interactions ($\gamma_{11} = .34$, $p < .01$). ICP was not significantly related to intimacy of opposite- or mixed-sex interactions ($\gamma_{21} = .00$, $\gamma_{31} = .10$, respectively). When interpreting regression coefficients, it is common practice to estimate predicted scores for people ± 1 *SD* on independent variables. ICP was a factor score (*SD* = 1.0), and so to estimate predicted values for people ± 1 *SD* on ICP, the (coefficients were multiplied by 1.0

and –1.0. The estimated same-sex intimacy for a person +1 *SD* on ICP was 4.11 (3.77 + [1*.34]), whereas for a person –1 *SD* on ICP it was 3.43 (3.77 + [–1 *.34]).

Differences in the strength of these relationships can also be compared statistically using chi-squared tests of fixed effects. For this particular example, the relationship between ICP and intimacy in same-sex interactions was significantly different from relationships between ICP and intimacy in opposite- and mixed-sex interactions (*p*s < .01). The ICP–intimacy relationships for opposite- and mixed-sex interactions were not significantly different.

It is important to note that this ability to compare within-person relationships constitutes an important advantage of HLM over comparable OLS techniques. For example, at the between-subject level it is not appropriate to use a Fisher's *r–Z* transformation to compare correlated correlations (e.g., the strength of relationship between ICP and intimacy in same- and opposite-sex interaction). Although it is possible to compare correlated correlations within an OLS framework (e.g., Cohen & Cohen, 1983, pp. 479–480), such procedures are much more cumbersome than the multiparameter test available in HLM. Moreover, such procedures do not take into account the multilevel nature of interaction diary data, and may therefore provide inaccurate estimates of relationships.

Contrast codes. Differences among types of interactions can also be examined using contrast codes. The important difference between contrast and dummy-coded analyses is that contrast codes model a *difference score* at the interaction level. That is, the dependent variable is now a difference score. For example, the *difference* in intimacy in same-sex versus opposite- and mixed-sex interactions could be examined with a single interaction-level variable consisting of a contrast code that assigned 2 to same-sex interactions and –1 to opposite- and mixed-sex interactions.

$$y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{S-OM}) + r_{ij}.$$

Additional analyses of the Nezlek and Leary data carried out for this article found that the mean S-OM contrast was significantly different from 0 ($\gamma_{10} = -.30$, $p < .01$). On average, the difference between intimacy in same-sex interactions versus opposite- and mixed-sex interactions was –.30.

At the purely between-person level, contrasts of differences in mean reactions to different types of interactions can also be analyzed using the dummy-coded scheme described earlier, although the two analyses can produce slightly different estimates because dummy codes analyze differences of means, whereas contrast codes analyze means of differences. For example, the S-OM contrast estimated from the means in the dummy-coded analysis is .37. Differences between the two procedures are due to individual differences in the distributions of different types of interactions. At present, it is not clear when dummy and contrast codes provide more or less accurate estimates. Although the two types of analyses typically produce similar results (i.e., similar estimates of effect sizes and similar significance levels for such differences), analysts are advised to be aware of possible differences produced by the two types of analyses.

An important advantage of contrast codes is the ability to model individual differences in difference scores. This cannot be done with dummy codes because dummy codes do not represent interaction-level contrasts. For example, differences between men and women in the S-OM contrast could be examined by modeling the S-OM coefficient at level 2. The level 2 variable, Sex, is also a contrast-coded variable (–1 = *men*, 1 = *women*), entered uncentered.

$$\text{Intercept:}\quad \beta_{0j} = \gamma_{00} + \gamma_{10}\,(\text{Sex}) + u_{0j}$$

$$\text{Contrast:}\quad \beta_{1j} = \gamma_{10} + \gamma_{11}\,(\text{Sex}) + u_{2j}.$$

Analyses of the Nezlek and Leary data done for this article found a significant difference between men and women in the S-OM contrast. This difference can be understood by estimating S-OM contrast scores for men and women. The intercept for the contrast equation (γ_{10}) was –.34, and the coefficient for sex (γ_{11}) was .31. This means that the estimated mean contrast for men was –.65 (–.34 + [–1 *.31]), whereas the estimated mean contrast for women was –.03 (–.34 + [1 *.31]). Consistent with previous research, intimacy varied less as a function of the sexual composition of interactions for women than it did for men.

Note that the intercept for the contrast equation was slightly different in this model than it was in the previous analysis in which the contrast was not modeled as a function of participant sex. This is because uncentered contrast-coded predictors produce intercepts that are corrected for predictor differences in the dependent variable (e.g., sex differences in the contrast). When Sex was entered centered at level 2, the intercept was the same as it was in the original model when Sex was not included. If a sample had the same number of men and women, or if there was no difference between the sexes, including Sex uncentered and centered would produce the same intercept, and this intercept would be the same as if Sex was not included as a predictor. Adjustments such as these are similar to those found in OLS regression except that, in most regression analyses, predictors are centered by default. Nevertheless, analysts whose primary experience is with balanced ANOVA designs might want to pay particular attention to how including different predictors (and how they are centered) influences results at all levels of analysis.

Differences between dummy and contrast-coded analyses highlight the flexibility and power of general linear modeling within the multilevel framework, and they highlight the importance of being particularly mindful about exactly what a coefficient represents. In single-level multiple regression, the distinctions between the two types of analyses are relatively if not completely unimportant – researchers are usually interested only in the significance of a coefficient per se. In multilevel analyses, the situation is different, however, because the slopes and intercepts from one level of analysis become dependent measures at the next level. Through the careful selection of dummy and contrast-coded predictors, analysts can isolate and model very specific measures; however, this flexibility is accompanied by an equally powerful need to keep in mind exactly what each coefficient and intercept represents when specifying models and interpreting results. To avoid confusion, analysts are strongly encouraged to interpret their results by generating predicted values using the coefficients estimated by their models, keeping in mind how predictors are centered.

Analyzing differences in slopes across mutually exclusive categories

Analyses of how slopes (e.g., relationships between two ratings of an interaction) vary across types of interactions are somewhat more complicated than analyses of means. There are two ways of doing this when interactions are classified using mutually exclusive types. The first involves adding a level of nesting, and such analyses will be illustrated using data collected by Sullivan, Nezlek, and Jackson (2001). In this study, 126 undergraduates maintained a variant of the RIR for two weeks. For each interaction, they rated (using 1–7 scales) how enjoyable and intimate the interaction was, how much social support they needed, received, and provided, and how much support their interaction partners needed. Similar to the Nezlek and Leary study, interactions were classified as being either same-, opposite-, or mixed-sex.

In these analyses, the dependent measure was social support received and the predictor was social support provided (group-mean centered). In essence, this analysis examined the reciprocity of social support. In this three-level model, interactions (subscripted i) were nested within types (subscripted j) and types were nested within people (subscripted k). The level 2 model consisted of dummy-coded variables (uncentered) representing same-, opposite-, and mixed-sex interactions respectively.

Interaction (level 1) $y_{ijk} = \pi_{0jk} + \pi_{1jk}$ (Support Provided) $+ e_{ijk}$

Type (level 2) $\pi_{0jk} = \beta_{01k}$ (Same) $+ \beta_{02k}$ (Opposite) $+ \beta_{03k}$ (Mixed) $+ r_{0jk}$

$\pi_{1jk} = \beta_{11k}$ (Same) $+ \beta_{12k}$ (Opposite) $+ \beta_{13k}$ (Mixed) $+ r_{1jk}$

Person (level 3) $\beta_{01k} = \gamma_{010} + u_{01k}$

$\beta_{02k} = \gamma_{020} + u_{02k}$

$\beta_{03k} = \gamma_{030} + u_{03k}$

$\beta_{11k} = \gamma_{110} + u_{11k}$

$\beta_{12k} = \gamma_{120} + u_{12k}$

$\beta_{13k} = \gamma_{130} + u_{13k}.$

At the interaction level, the relationship between support received and provided was represented by the π_{1jk} coefficient. This coefficient was then estimated separately for same-, opposite-, and mixed-sex interactions at level 2, and, in turn, the mean slopes for the three types of interactions were estimated at level 3.

The mean slopes for same-, opposite-, and mixed-sex interactions (the γ_{110}, γ_{120}, and γ_{130} coefficients) were .39, .45, and .47, which were all significantly different from 0 (ps < .01). These slopes are interpreted as follows. For each 1.0 increase in support provided, support received increased .39 in same-sex interactions, .45 in opposite-sex interactions, and .47 in mixed-sex interactions. A contrast (–2, 1, 1) of these slopes using the chi-squared-based multi-parameter test of fixed effects found that the same-sex slope was significantly less than the opposite- and mixed-sex slopes ($p < .05$). Finally, it is critical to note that because support provided was group-mean centered, these analyses controlled for individual differences and differences across the three types of interactions in support provided.

Differences in slopes across types of interactions can also be examined using a two-level model that includes within-person interaction terms, and such a procedure may be particularly convenient when only two categories are being compared (e.g., dyads versus nondyads). Following the advice of Aiken and West (1991), the continuous predictor variable should be centered. In the multilevel case, this means subtracting each person's mean score on the predictor from the score for each interaction (do not center around the grand mean). This centered score should then be multiplied by either a dummy or contrast-coded variable representing type of interaction (keeping in mind the implications such coding has for the interpretation of the intercept). The dependent measure is then modeled as a function of the type of interaction (uncentered), the predictor (group-mean centered), and the interaction term (uncentered).

This type of analysis is illustrated from data collected by Sullivan et al. (2001). The dependent measure is support received, the predictor variable is support needed, and the analysis examines how this relationship varies between dyadic and nondyadic interactions. The following level 1 model was used:

$$y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Dyad}) + \beta_{2j}\,(\text{Needed}) + \beta_{3j}\,(\text{Dyad–Needed}) + r_{ij}.$$

The significance test of the mean Dyad–Needed coefficient (γ_{13}, done at level 2 as in the previous examples) represents the test of whether the mean interaction term is 0 across all participants. The analyses of the Sullivan et al. data produced the following coefficients:

$$y_{ij} = 2.65 + .74\,(\text{Dyad}) + .08(\text{Needed}) + .03(\text{Dyad–Needed}) + r_{ij}.$$

All three coefficients were significantly different from 0 (ps < .05), and the nature of the interaction can be understood through predicted scores. Typically, interactions involving continuous variables are interpreted using scores ± 1 *SD* from the mean. For present purposes, high and low support needed were defined as ± 1 *SD* on support needed, and the interaction was examined using predicted scores for high and low need dyads and for high and low need nondyadic interactions.

The within-person *SD* for support needed was obtained from an unconditional analysis of support needed, the estimated level 1 *SD*, which was 1.50. This meant that 1 *SD* change on support needed corresponded to .12

TABLE 2
Predicted scores for support received as a function of support needed and dyadic nature of interactions

		Received	Intercept	Dyad	Need	Interaction
Coefficient Dyad			2.65	+.74	.08	.03
	High Need	3.56	2.65	+.74	+.12	+.05
	Low Need	3.22	2.65	+.74	–.12	–.05
Non-Dyad	High Need	1.98	2.65	–.74	+.12	–.05
	Low Need	1.84	2.65	–.74	–.12	+.05

change in support received (.08 * 1.50). On average, for interactions that were 1.50 points above a person's mean support needed, support received was .12 above the person's mean support received; for interactions that were 1.50 points below a person's mean support needed, support received was .12 below the person's mean support received. These deviation scores were used because support needed was group-mean centered. The contribution of the interaction term was estimated by multiplying the coefficient for the interaction term (.03) by the product of dyad and need scores. For example, the contribution of the interaction for high need dyads was +.05 = .03 * (1 * 1.50), whereas the contribution of the interaction for low need dyads was –.05 = .03 * (1 * –1.50).

Predicted support received scores for four types of interactions are presented in Table 2. These means describe the large main effect for dyad, and the main effect for need (i.e., the difference in support received between low and high need interactions). They also indicate that the difference between high and low need interactions in support received was larger for dyadic interactions (.34 = 3.56–3.22) than it was for nondyadic interactions (.14 = 1.98–1.84). It is important to note that cross-multiplied centered scores can also be used to examine interactions between two continuous measures (e.g., Nezlek & Plesko, 2003).

Analyzing differences across interactions for nonexclusive categories

Although relationships researchers may be interested in mutually exclusive systems (e.g., dyads versus nondyads), there is considerable interest in understanding how interactions differ as a function of what are not necessarily nonexclusive typologies. A good example of a nonmutually exclusive system would be describing interactions on the basis of the relationships participants had with the others present at an interaction. Although dyads can be defined by mutually exclusive categories because only one other person is present, approximately 50% of interactions involve more than 1 other person. Such nondyadic interactions may include multiple relational partners (e.g., both friends and romantic partners), making it impossible to classify the interaction neatly into a particular category.

Comparing interactions using a system that does not consist of mutually exclusive categories is somewhat more complicated than when categories are mutually exclusive. Interactions cannot be assigned to only one of a set of categories, and this can make the types of analyses described earlier (with either two- or three-level models) problematic. One solution to such problems consists of creating a mutually exclusive system that is then analyzed using the techniques described earlier. For example, assume a researcher is primarily interested in interactions with romantic partners. Interactions could be classified as involving romantic partners only, involving romantic partners and other people, or not involving romantic partners. This would comprise a mutually exclusive system in that a particular interaction falls into only one of these categories.

The advantage of this approach is the flexibility afforded by the use of mutually exclusive systems. For example, means and slopes can be compared directly and unambiguously. A disadvantage of this approach is that classifying an interaction based on the absence of a specific partner may mean that interactions with meaningfully distinct relational partners are not distinguished. For example, interactions that did not involve a romantic partner could involve friends, strangers, family members, and so forth, and it might not be appropriate to consider all these relational partners as functionally equivalent (i.e., not a romantic partner).

This difficulty can be remedied by adding categories that distinguish interactions with meaningfully distinct relational partners. For example, for collegians, the vast majority of interactions involve two types of relational partners, same-sex friends and romantic partners. Assuming this, interactions could be classified using a four-category system: romantic partners only, friends only, friends and romantic partners, and neither romantic partners nor friends. Deciding if interactions can be classified using a mutually exclusive system requires having a sense of how important it is to distinguish different types of interactions and knowing the distribution of interactions of different types. Although it may be theoretically interesting to distinguish certain types of interactions (e.g., between students and professors), if a type of interaction is infrequent (e.g., between students and professors) it will not be possible to model this type of interaction as a distinct category. One cannot model what does not occur.

Contrast codes can also be used to compare mean ratings of interactions when a nonmutually exclusive categorical system is of interest. For example, the relative influence of romantic partners, friends, and family members could be examined with the following model:

$$y_{ij} = \beta_{0j} + \beta_{1j}\ (\text{Romantic}) + \beta_{2j}\ (\text{Friend}) + \beta_{3j}\ (\text{Family}) + r_{ij}.$$

The data collected by Nezlek et al. (2000) were coded for the presence or absence of each of these relational partners (–1 = *no*, 1 = *yes*), and these predictors were entered uncentered. The dependent measure was enjoyment, and the analyses found a significant effect for all three types of relational partners. The intercept was 5.50, and the ($_{10}$, ($_{20}$, and ($_{30}$ coefficients were .63, .72, and .64, respectively, which were all significantly different from 0 (ps < .01). This meant that on average, the presence of a romantic partner was associated with 1.26 difference in enjoyment. The predicted enjoyment for interactions without a romantic partner was 4.87 [5.50 + (–1 *.63)], whereas it was 6.13 [5.50 + (1 *.63)] for interactions with a romantic partner.

It is *absolutely critical* to recognize that this coefficient represents the difference between interactions with and without a romantic partner *controlling* for the presence or absence of friends and family members. When the romantic partner effect was examined in isolation (i.e., when the predictors for friends and family members were not included in the level 1 model), the coefficient for the presence or absence of a romantic partner was .32, approximately half of what it was in the analysis including family and friends. The difference between the two estimates is due to the fact that interactions that did not involve a romantic partner may have involved friends or family members, relational partners whose presence was associated with increased enjoyment. When the friend and family contrasts were not in the model, the estimated mean for nonromantic interactions was not adjusted for the presence of family members and friends, leading to an increase in estimated mean enjoyment for nonromantic interactions. When the friend and family contrasts were in the model, the estimated enjoyment for interactions that did not involve romantic partners was adjusted for the presence of family members and friends.

The differences in the results of these two analyses highlight the subtleties of analyses involving the presence of relational partners. Determining which types of relational partners need to be included in a model needs to be guided by both theoretical and empirical concerns. It might be theoretically important to control for the presence or absence of certain types of relational partners simultaneously, whereas, empirically, the presence or absence of some relational partners may have impacts that were not anticipated. Researchers need to anticipate such possibilities as they design studies, with the understanding that more finely detailed categorical systems can always be collapsed, whereas data collection systems relying on relatively coarse distinctions cannot be expanded.

Evaluating differences in slopes across different types of interactions using nonexclusive systems can also be complex, particularly if a mutually exclusive system cannot be devised. Without a mutually exclusive system, three-level models with type of interaction dummy coded at level 2 cannot be used because interactions cannot be nested within type of interaction because a particular interaction can be classified as more one type. This leaves only the two-level option with cross-multiplied terms, similar to the analyses of dyadic versus nondyadic interactions discussed previously.

Although this two-level option solves the immediate problem (i.e., the slopes for two types of interactions can be compared), it leaves open the possibility that meaningfully distinct types of interactions may be combined when a dichotomy is based on the presence or absence of a specific relational partner. For example, a comparison of slopes for interactions with and without romantic partners potentially combines into the ‘not’ category interactions with various relational partners that might be worth distinguishing. Analysts who are concerned about the possible overlap between such categories can include terms representing interactions between the predictor and each category. This entails adding terms representing the 'main effects' for each type of relational partner (e.g., a contrast-coded variable) and the cross-product of each of these terms with the predictor. Although there is no theoretical limit to the number of variables that can be included at any level of analysis, analysts should be conservative in adding variables. As discussed later, adding a variable to a multilevel analysis can require estimating numerous additional parameters, taxing what has been described as the 'carrying capacity' of a data structure.

The foregoing was not intended to be an exhaustive description of all the different ways that reactions to interactions can be analyzed. Rather, analyses were described that examine relationships that should be of interest to a broad range of relationship researchers. Through the use of various combinations of contrast and effect codes and centering options, virtually any type of interaction can be isolated or compared with any other type – the only limit on such combinations is the imagination of the analyst.

Interaction quantity

Unlike ratings of interactions, which can be considered primarily, if not exclusively as an interaction-level phenomenon, quantity of interaction can be conceptualized in terms of various bases. That is, interaction quantity can be thought of in terms of absolute quantity – How many interactions do people have and how much time do they spend in interaction? Interaction quantity can be also thought of in terms of relative distribution, as a characteristic of an interaction – What percentage of interactions involve romantic partners? Interaction quantity can also be thought of in terms of social networks – With how many different people does someone interact each day or during a typical week? Procedures that examine quantity of interaction in terms of each of these bases are described.

Absolute quantity of interaction

Before conducting analyses of absolute quantity of interaction, analysts need to specify the temporal and interaction units that will serve as the bases for analysis. That is, some form of summary measure needs to be calculated over some period representing the quantity of specific types of interaction. For example, the day is a powerful organizing temporal unit for people, and so the number of interactions that occurred each day or the time spent in interaction could be calculated. Other temporal units such as mornings and evenings could be used. The number of interactions and time spent in interaction could also be calculated separately for same-, opposite-, and mixed-sex interactions, for dyadic versus nondyadic interactions, interactions with romantic partners, and so forth.

Size of social network can be analyzed in two ways: (1) The number of different people with whom someone interacted each day, and (2) the number of different people with whom someone interacted over the course of a study. Either type of measure can be divided into categories such as same- or opposite sex. Number of different people per day can be analyzed using the techniques described below. Number of different people for an entire study is an individual (person-level) measure that can be analyzed using traditional OLS techniques.

Aside from the need to calculate summary measures, the critical difference between analyses of absolute quantity and analyses of interaction ratings is that for analyses of absolute quantity, temporal units (e.g., days), not interactions, are nested within people. The models used in these analyses are the same as those used in the analyses of ratings, however. The basic level 1 model is similar to that used in the analyses of interaction quality:

$$y_{ij} = \beta_{0j} + r_{ij}.$$

In such a model, y_{ij} is a measure of quantity of interaction for a temporal unit (subscripted i) for each participant (subscripted j), β_{0j} is a random coefficient (an intercept) representing the mean of y across temporal units, r_{ij} represents error, and the variance of r_{ij} represents the level 1 residual (error) variance. This model is accompanied by a level 2 model that is also similar to that used in analyses of interaction ratings:

$$\beta_{0j} = \gamma_{00} + u_{0j}.$$

In this model, γ_{00} represents the grand mean of the person-level means (β_{0j}s) from the level 1 model, u_{0j} represents the error of β_{0j}, and the variance of u_{0j} constitutes the level 2 residual (error) variance.

Such a scheme was used by Nezlek et al. (2000) to examine relationships between depression and interaction quantity. Across the 43 participants in the study, data were collected for 574 days. For each person, for each day the diary was maintained, summary scores were calculated representing the time people spent per day in interaction and how many interactions they had. Similar to the analyses of interaction quality, diagnosis was represented at level 2 with a contrast-coded variable (–1 = *depressed*, 1 = *not depressed*). The analysis of time spent per day in interaction produced an intercept (γ_{00}) of 243 and a diagnosis coefficient (γ_{01}) of 21. The diagnosis coefficient was not significantly different from 0, indicating that the time depressed participants spent each day in interaction [222 = 243 + (–1 * 21)] was not significantly less than the number for the nondepressed [264 = 243 + (1 * 21)]. A similar analysis of interactions per day also revealed no significant differences between the two groups (3.6 versus 4.6).

Additional confidence in the conclusion that the quantity of interaction of these two groups did not differ was found when cohabiting status was included as a level 2 predictor. For example, when time per day spent in interaction was modeled as a function of cohabitation status and diagnosis, the coefficient for diagnosis was reduced to 3, corresponding to a difference of 6 minutes per day between the two groups. A similar analysis of interactions per day also produced a smaller coefficient for diagnosis (.41 versus

.52). Such analyses highlight the importance of including additional variables in models to eliminate possible confounded effects. For example, in the Nezlek et al. study, diagnosis and cohabitation were confounded. Consistent with other research, fewer depressed people were cohabiting than the nondepressed, creating the possibility that any differences in contact between the two groups might reflect the availability of a cohabiting partner rather than the effects of depression per se.

It is important to note that such summary measures can be calculated to represent various types of interactions. In studies of collegians, it has been useful to distinguish same-, opposite-, and mixed-sex interactions, and separate summary measures can be calculated representing quantity of these types of interactions. Similarly, summary measures can be calculated to represent quantity of interaction with different relational partners. When analyzing summary measures representing different types of interactions, analysts need to be mindful of whether or not the categorical system was mutually exclusive, because when nonexclusive systems are used, a specific interaction can contribute to multiple summary measures.

Distribution of interaction. Another way of understanding quantity of interaction is by examining the relative quantity of interactions, operationalized in terms of percent of interaction. Analyses of percents and other measures that are by definition not normally distributed (e.g., categorical measures, highly skewed count data) rely on the same logic as analyses of normally distributed measures, although they use slightly different techniques. Different techniques are necessary because the distributions of such measures violate the assumption of the independence of means and variances. For example, the variance of a binomial is npq, where n = number of observations, p = the probability of the more common outcome, and $q = 1 - p$.

To analyze categorical and count (poisson) outcomes, HLM uses techniques that rely on log-odds. For example, the percentage of interactions that were dyads could be analyzed with a two-level model in which interactions were nested within participants, and the dependent measure was a dummy coded variable representing if an interaction was a dyad or not. In such a case, the interaction-level model would be a Bernoulli model with $n = 1$.

$$\text{Prob}(y = 1|\beta_{0j}) = \phi.$$

The coefficients from this model, the log-odds of a particular type of event occurring, can be analyzed at the person level using the same techniques as those used in previous analyses. Moreover, predictors can be added at levels 1 and 2 in the same way they can be added to analyses of linear outcomes. An example of such analyses can be found in Nezlek (1999), and a description of using HLM to analyze categorical and count outcomes can be found in Raudenbush et al. (2000, pp. 111–164). Analysts are advised to be particularly cautious when conducting and interpreting nonlinear analyses. Although the transformations are clearly described in Raudenbush et al.

(2000), producing predicted values, which are needed to understand the coefficients, can be quite vexing for those who are not familiar with working with log-odds.

Special topics

Missing data

Unlike OLS regression and ANOVA, HLM allows missing data of some types. For two-level models, missing data are allowed at level 1 but not at level 2. For three-level models, missing data are allowed at level 1 but not at levels 2 and 3. For nonlinear analyses, no missing data are allowed at any level. When creating the system file on which HLM relies to conduct analyses, analysts can specify pairwise or listwise deletion. Deciding which of these options to use needs to take into account the potential bias each form of deletion may introduce. For example, using listwise deletion, any missing datum will exclude a case (e.g., an interaction) from analysis. If many missing data are scattered throughout a data set, this may mean that many cases are excluded, leaving only fully complete cases, a subset that may or may not represent the full sample of interactions. With pairwise deletion, analyses involving variables for which there are many missing responses may provide inaccurate parameter estimates because these estimates are based on a potentially nonrepresentative sample.

There are no hard and fast rules for how to take into account missing data. The primary principle guiding such decisions should be the reason why data are missing. If there are few missing data and there is no pattern to them, how missing data are taken into account will not matter much. In contrast, if data are missing systematically (e.g., groups of individuals have not provided specific responses), how missing data are taken into account may matter.

In social interaction diary studies, systematically missing data can occur for reasons other than participant failure to provide responses. For example, some participants may not have interactions with certain relational partners (e.g., romantic partners). Models (such as the one below) that include terms representing the presence or absence of such partners may provide misleading estimates of the effects for the presence or absence of these partners, and misleading estimates of other effects (because MRCM relies on covariance matrices to estimate effects).

$$y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Romantic}) + \beta_{2j}\,(\text{Friend}) + \beta_{3j}\,(\text{Family}) + r_{ij}.$$

At present, the implications of different ways of dealing with such situations are not well understood, and different options may be appropriate depending on the questions at hand. For example, Nezlek et al. (2000) were interested in how relationships between depression and reactions to interactions varied across interactions with different relational partners (romantic partners, friends, co-workers, family, and strangers). Many participants did not have interactions with all these different relational partners, leaving various options. Analyses could have been done using only the subset of participants who had interactions with all partners, or analyses could have been done on the full sample with the understanding that all participants (regardless of whether they had interactions with certain relational partners) would contribute to the estimate of effects for all relational partners.

To avoid the problems inherent in each of these strategies, Nezlek et al. (2000) analyzed different subsets of participants (determined on the basis of whether certain relational partners were present in a diary) with models that used dummy-coded variables to estimate ratings of interactions with specific partners. This allowed the greatest number of interactions to be included for the analysis of each type of relational partner, while ensuring that each analysis included only participants who had interactions with each type of partner. Analysts interested in other questions (e.g., comparing interactions with different relational partners) may need to pursue other strategies (e.g., contrast-coded analyses).

Until more formal studies have been done to determine how different patterns of systematically missing data influence parameter estimates, it is not possible to provide definitive advice regarding how to analyze such data structures. Analysts need to be mindful of the fact that systematically missing data may lead to biased estimates and how to avoid such biases is likely to vary as a function of the specific situation at hand. Analysts who are particularly concerned about such issues may want to conduct different types of analyses and compare the results obtained from each.

Modeling random error

An important feature of MRCM analyses is that they estimate both a fixed and a random error term for each coefficient in a model. Fixed terms represent estimates of population means for coefficients, whereas random error terms represent estimates of the error variance of coefficients. Recall that MRCM analyses separate random and true variance. Moreover, tests of fixed effects are appropriate for the overwhelming majority of the questions personal relationships researchers are likely to pose. Nonetheless, random error terms are an important part of any MRCM analysis, and analysts need to be mindful of how they specify random error in their models. For example, the inclusion (or deletion) of a random error term for a specific coefficient can change the results of the significance test of the fixed effect of that coefficient, and it is not possible to predict which change will occur. Sometimes deleting a random error term will make a significant fixed effect nonsignificant, and sometimes a nonsignificant fixed effect will become significant. Moreover, because MRCM analyses rely on covariance matrices to estimate parameters, including or deleting individual random error terms can change the estimation of other parameters, both fixed and random.

A coefficient is referred to as 'random' if a random error term is estimated for it and is referred to as 'fixed' if no random error term is estimated. For most social interaction diary research, effects should be modeled as random. This reflects the fact that, typically, the interactions described in a study have been randomly sampled from the population of interactions a person might have. As discussed by Nezlek (2001), there are certain times, however, when effects can or should be fixed: (1) If a specific random error term cannot be estimated reliably (i.e., it is not significantly different from 0); (2) If a model has trouble converging (i.e., not converging within 100–200 iterations) due to difficulties estimating error variances and covariances; or (3) For theoretical reasons such as the possibility that the interactions described in a study have not been randomly sampled from a population but represent a population (e.g., interactions between a husband and wife during the first week of a first marriage). Regardless of whether effects are modeled as random or fixed, analysts need to describe how error was modeled, and, for effects that are modeled as fixed, they should provide some rationale for why the effect was modeled as fixed rather than as random.

A common misunderstanding among some analysts new to random effects models is the belief that if the random error term associated with a coefficient is not significant, there is no variation in the fixed effect associated with that coefficient. Although this may be the case, it may also be the case that random error term cannot be estimated reliably for other reasons. That is, the nature of the variability is such that random (error) and true variance cannot be estimated separately and reliably, and there may be meaningful variation in the fixed effect for a coefficient even when the random error term cannot be estimated reliably. Within the nomenclature of multilevel modeling, such coefficients are referred to as ‘nonrandomly varying’ because the coefficients vary but have no random error term associated with them. Analysts need to recognize that tests of random terms do not constitute tests of hypotheses that there is no variance in the coefficient. Rather, tests of random error terms indicate only if a random error term can be reliably estimated for a coefficient (i.e., can true and error variance be reliably separated).

Model building

Building a model (making a final determination of the relationships among the variables for a sample) within MRCM is somewhat more complicated than in OLS single-level analyses. Part of this complexity is because, within the multilevel framework, models can be built at any level in relative isolation and at all levels simultaneously. For example, with a two-level data structure in which interactions are nested within people, models can be built to describe interaction-level phenomena (e.g., the covariance between two measures), person-level phenomena (e.g., the relationship between a mean rating of interaction and person-level characteristics such as personality traits), and cross-level effects (e.g., does an interaction-level covariance vary as a function of person-level characteristics). Each of these types of models has been presented previously.

Based on their experience with OLS single-level analyses, the tendency of many analysts is to use what are called ‘backward stepping’ procedures. Initial models include many variables, and variables that are not significant are excluded. One advantage of this procedure is that it takes into account the covariances among multiple predictors. Although potentially appealing because of this, such backward stepping procedures are generally not appropriate for MRCM. The difficulty with including large numbers of predictors is that, because MRCM estimates error covariance matrices, adding a single term to a model can increase dramatically the number of parameters a model is estimating. For example, assume a level 1 model with two predictors. Adding a third variable adds four parameters, the fixed effect for the new predictor, the error variance of the new predictor, and the covariance between the new predictor and each of the existing predictors. As variables and parameters are added, models may begin to tax the 'carrying capacity' of a data set – the number of parameters a data set can estimate. When analyzing social interaction diary data, analysts are advised to use 'forward stepping' procedures, starting with relatively simple models that test specific hypotheses or predictions and then determining how such relationships vary across situations or are affected by other measures.

Models with many predictors also increase the likelihood of experiencing problems due to collinearity. Similar to single-level OLS analyses, collinearity at either level 1 or 2 may make estimates less stable (i.e., individual coefficients may vary greatly as a function of the specific variables included in a model). Moreover, collinearity at level 1 may make it difficult to estimate parameters, which may make it more difficult for models to converge. Analysts are advised to avoid models in which predictors are highly collinear.

Within a multilevel context, another important consideration is the order in which models are built. For a social interaction diary study, this refers to the question: Does one finalize the interaction- or person-level model first? Unfortunately, there are no clear-cut answers to this question, although it is possible to suggest some guidelines. Perhaps the most important consideration is the research question at hand: Is the emphasis on person- or interaction-level relationships? If person-level relationships are of primary importance, an analyst may start with a simple (unconditional) interaction-level model and elaborate the model at the person level. For example, what are the relationships between mean intimacy in interaction and various individual differences? In many respects, such analyses are very much like OLS regressions in which mean intimacy is the dependent measure and individual difference variables are the predictors.

Although such broadly focused analyses can be good starting points, the true value of social interaction diary studies lies in the opportunities they provide to examine different types of interactions. Such analyses can focus on differences among interactions per se (e.g., how do interactions with friends and lovers differ?) and on how relationships between person-level measures and interaction measures vary across different types of interactions (e.g., is the relationship between extraversion and satisfaction different for interactions with friends and lovers?). Such analyses require elaborating level 1 (interaction-level) models, perhaps using the techniques discussed previously.

There are numerous ways to classify interactions and to conduct analyses representing different classificatory systems, and, certainly, sound theory and findings from existing research provide the best bases for deciding how to differentiate interactions. Nonetheless, regardless of how they differentiate interactions, scholars need to be sensitive to possible differences across different types of interactions. Moreover, when examining relationships between interaction- and person-level constructs, analysts should consider a sort of iterative strategy in which broadly focused models at one level are combined with more specifically focused analyses at the other. For example, Nezlek et al. (2000) started with a relatively simple analysis examining relationships between depression and reactions to interactions. The person-level model was elaborated by including terms for participant sex, SES, etc., and the interaction-level model was elaborated by examining how relationships between depression and reactions varied across interactions with different relational partners. Nezlek and Leary (2002) used a similar process, although their level 1 model differentiated interactions using a different approach, the sex composition of the event (same-, opposite, or mixed-sex). Final models may be elaborated at both levels, but the process of forward stepping and discovering conditions in which simple relationships are maximized, minimized, or reversed can provide a rich understanding of the phenomena under investigation.

Standardizing variables

As mentioned previously, MRCM estimates and analyses unstandardized, as opposed to standardized, coefficients. This is primarily because unstandardized covariance matrices have more information (i.e., means and variances for individual variables) than standardized matrices and because MRCM was developed by educational and sociological researchers for whom real metrics (e.g., reading levels and dollars) are more meaningful than they are for many relationship researchers. In most social interaction diary studies, choices of metrics are relatively arbitrary (e.g., 7- versus 9-point response scales), and particular metrics typically have no inherent value. Such a relative lack of concern for scale characteristics generally poses few or no problems when conducting OLS analyses of interaction diary data because most analysts focus on standardized coefficients.

In contrast, in MRCM, scale metrics do matter, and, in some cases, changing metrics can change the results of significance tests. Similar to OLS analyses using standardized coefficients, the results of MRCM analyses of individual fixed effects (i.e., is an effect different from 0?) are invariant under linear transformation. Scale metrics (and resulting differences in variances) do matter, however, when coefficients are compared. For example, recall that the interaction-level model in which one rating (satisfaction) was modeled as a function of intimacy and influence:

$$y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Intimacy}) + \beta_{1j}\,(\text{Influence}) + r_{ij}.$$

Any difference in the variances between intimacy and influence contributes to the test of the difference between the mean slopes for intimacy and influence. In Nezlek et al. (2000), all three ratings used the same metric (a 1–9 scale) and had similar variances, so differences in scale metrics did not contribute to the test of the two mean slopes. If intimacy and influence had been measured using different scales, the results of significance tests of the mean slopes would have been different because the slopes themselves would have been different.

Metrics of level 2 (person-level) variables can also contribute to tests comparing the strength of relationships between level 1 coefficients and level 2 variables. For example, assume an interaction-level model that estimates mean intimacy:

$$y_{ij} = \beta_{0j} + r_{ij}.$$

The hypothesis of interest concerns the relative strength of the relationships between intimacy and extraversion (γ_{10}) and between intimacy and a rating of physical attractiveness (γ_{20}).

$$\beta_{0j} = \gamma_{00} + \gamma_{10}\,(\text{Extraversion}) + \gamma_{10}\,(\text{Attractiveness}) + u_{0j}.$$

Unlike OLS tests of standardized coefficients, the results of the test of the difference between these two coefficients depend on the variances of the two measures. For example, the results of analyses using extraversion measured on 1–5 scale and attractiveness measured on a 1–100 scale will differ from the results of analyses using attractiveness ratings divided by 20.

The influence of differences in variances on the results of significance tests can be eliminated by standardizing variables. At level 2 (or the highest level in a model), standardization is relatively straightforward; individual scores can be transformed into *z*-scores. At levels other than the highest level in a model, standardization is not so straightforward. If observations are standardized in reference to the full sample (i.e., ignoring person-level differences in means), the resulting *z*-scores confound interaction- and person-level variances – something diametrically opposed to the rationale for conducting a multilevel analysis. If observations are standardized in reference to each individual's data, mean differences are eliminated (everyone has a mean of 0), and important information is lost.

Researchers need to anticipate such problems and design studies in ways that avoid such problems. For example, use the same scale for different measures. If this is not possible, observations can be divided or multiplied to change variances, while still retaining mean differences. Regardless of how data are transformed, analysts need to be particularly mindful of the implications transformations have for tests of coefficients in multilevel analyses. See Nezlek (2001) for additional discussion of this issue.

Evaluating the strength of relationships

Another important difference between OLS and MRCM analyses is how effect sizes are evaluated. In both types of analysis, effect sizes can be described in terms of the sizes of coefficients: For every 1.0 change in *x*, how much does *y* change? Another way to conceptualize effect sizes is in terms of variance accounted for. In OLS analyses, effect sizes and tests of significance are both based on the sum of squares for an effect, and significance tests are based on changes in variance accounted for. In contrast, in MRCM analyses, fixed and random effects are estimated separately. In MRCM analyses, it is possible to add a variable to a model that has a statistically significant fixed effect, while the error variance remains unchanged.

Having noted this, shared variances can be estimated in MRCM analyses, albeit with some caution when multiple predictors are involved (Kreft & de Leeuw, 1998, pp. 115–119). To estimate effect sizes in terms of shared variance, residual errors from different models are compared. For example, to estimate the interaction-level variance shared by satisfaction and intimacy the level 1 residual variance from a model in which satisfaction is not modeled as a function of intimacy can be compared to the residual variance from a level 1 model in which satisfaction is modeled as a function of intimacy. In the Nezlek et al. (2000) study, these two variances were 3.40 and 2.39, a reduction of 30%, corresponding to a correlation of .54 (the square root of .30). Similarly, the variance of a level 1 coefficient (intercept or slope) accounted for by a level 2 variable can be estimated by comparing the level 2 residual variance from a model in which the level 1 coefficient is not modeled at level 2 to the residual variance from a level 2 model in which the coefficient is modeled. For example, in the Nezlek et al. study, the variance of the intercept of satisfaction was 2.24 without depression at level 2, and it was 1.42 when depression was included at level 2, a reduction of 37%, corresponding to a correlation of .60.

Software options

Conducting MRCM analyses requires specialized software. Although numerous programs can perform MRCM analyses, at present, three programs seem to stand out for the types of analyses discussed in this article: HLM (Raudenbush et al., 2000), Mln (Rabash, Yang, Woodhouse, & Goldstein, 1995), and SAS PROC MIXED (Littell et al., 1996). Of these, HLM and Mln were created specifically to do the types of multilevel analyses described in this article, whereas PROC MIXED was designed as a general purpose modeling program that can be used to perform multilevel analyses (e.g., Singer, 1998). It is well beyond the scope of this article to compare these three options systematically, although a few brief comments may help readers decide which option (if any) is best for them.

Of the three (HLM, Mln, and SAS), HLM is probably the most user-friendly. Models are built via a graphical interface, and it is easy to change the specifications for a model (centering, specifications of error terms, etc.). Moreover, the modeling conventions (e.g., nomenclature) used in the program correspond to the conventions used by Bryk and Raudenbush (1992), a popular text for multilevel analysis. Possible shortcomings are some idiosyncrasies in file preparation (e.g., identifying fields need to be alphanumeric characters) and the fact that the program has no data transformation capabilities. In Mln, models are also built via a graphical interface (which may not be quite as accessible as that used by HLM), although

Mln has two advantages over HLM, data transformation capability and more flexibility in specifying covariance structures (something that is likely to be important only to relatively advanced users).

As part of the larger SAS system, SAS PROC MIXED can take advantage of the features of the SAS system (e.g., data preparation and transformation), which are quite powerful. However, SAS was not designed specifically to conduct multilevel random coefficient analyses of the sort described here, and, consequently, many users may experience difficulties accessing and understanding different instructional sets. Moreover, and particularly important for less experienced users, the output is not labeled as clearly and unambiguously as the output is for HLM and Mln. On balance, analysts who do not have any experience with multilevel analysis and who are not conversant with SAS and its numerous conventions may find it easier to use HLM or Mln than SAS. Experienced SAS users who have some knowledge of multilevel analysis may find it easier to stay with the more familiar SAS.

Multilevel analyses can also be conducted using different variants of structural equation modeling. Although such analyses are technically possible, more and more researchers are coming to believe that SEM is not as appropriate as MRCM for the types of multilevel data structures generated in social interaction diary studies (e.g., Schnabel, Little, & Baumert, 2000). For example, although it is technically possible to conduct SEM analyses in which each person is treated as a group and interactions are treated as observations, such analyses may be impractical when there are 50 or 100 groups (i.e., people) in a study.

Concluding remarks

Social interaction diary data provide both opportunities and challenges. By their nature, the data produced in social interaction diary studies are rich and complex, and frequently researchers are confronted with a bewildering array of options. Although MRCM per se will not indicate which analyses will test which hypotheses, MRCM constitutes a powerful analytic tool that provides an internally consistent framework for researchers. Moreover, this framework is quite flexible, allowing researchers to investigate a wide variety of hypotheses with precision. In addition to the increased accuracy provided by MRCM, widespread use of a common analytic framework should enhance our ability to compare and contrast results across studies, which in and of itself should enhance our understanding of social interaction.

There is a learning curve, whose steepness will vary a function of numerous factors, and, for some analysts, learning to use MRCM will not be easy. Nevertheless, as Kreft and de Leeuw (1998, p. 1) noted ‘Once you know that hierarchies exist, you see them everywhere.’ Researchers who learn how to use MRCM to analyze social interaction diaries will have more confidence in the conclusions they draw about social interaction and are likely to recognize how such techniques can be used to address other types of data structures with different concerns.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression and correlation analyses for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

DePaulo, B., Kashy, D. A., Kirkendol, S. A., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979–995.

Gable, S. L., & Reis, H. T. (1999). Now and then, them and us, this and that: Studying relationships across time, partner, context, and person. *Personal Relationships, 6*, 415–432.

Hodgins, H. S., Koestner, R., & Duncan, N. (1996). On the compatibility of autonomy and relatedness. *Personality and Social Psychology Bulletin, 22*, 227–237.

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. 1, pp. 233–265) New York: McGraw Hill.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park, CA: Sage.

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

Nezlek, J. B. (1995). Social construction, sex/sex similarity, and social interaction in close personal relationships. *Journal of Social and Personal Relationships, 12*, 503–520.

Nezlek, J. B. (1999). Body self-evaluation and day-to-day social interaction. *Journal of Personality, 67*, 793–817.

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event and interval contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin, 27*, 771–785.

Nezlek, J. B., & Derks, P. (2001). Use of humor as a coping mechanism, psychological adjustment, and social interaction. *Humor: International Journal of Humor Research, 14*, 395–413.

Nezlek, J. B., Hampton, C. A., & Shean, G. D. (2000). Clinical depression and everyday social interaction in a community sample. *Journal of Abnormal Psychology, 109*, 11–19.

Nezlek, J. B., Imbrie, M., & Shean, G. D. (1994). Depression and everyday social interaction. *Journal of Personality and Social Psychology, 67*, 1101–1111.

Nezlek, J. B., & Leary, M. R. (2002). Individual differences in self-presentational motives and daily social interaction. *Personality and Social Psychology Bulletin, 28*, 211–223.

Nezlek, J. B., & Plesko, R. M. (2003). Affect- and self-based models of relationships between daily events and daily well-being. *Personality and Social Psychology Bulletin*.

Nezlek, J. B., Richardson, D. S., Green, L. R., & Schatten-Jones, E.C. (2002). Psychological well-being and day-to-day social interaction among older adults. *Personal Relationships, 9*, 57–71.

Nezlek, J. B., & Wheeler, L. (1984). RIRAP: Rochester Interaction Record Analysis Package. *Psychological Documents, 14*, 6. (Ms. No. 2610).

Rabash, J., Yang, M., Woodhouse, G., & Goldstein, H. (1995). *MLn: Command reference guide*. London: Institute of Education.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2000). *HLM*. Chicago: Scientific Software.

Schnabel, K. U., Little, T. D., & Baumert, J. (2000). Modeling longitudinal and multilevel data. In T. D. Little, U. K. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 9–14). Mahwah, NJ: Erlbaum.

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., & Bosker, R. (1999). *Multilevel analysis*. 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.

Sullivan, L. A., Nezlek, J. B., & Jackson, L. A. (2001). *A multilevel analysis of providing, receiving, and exchanging social support in everyday social interaction*. Unpublished manuscript, College of William & Mary, Williamsburg VA.

Wheeler, L., & Nezlek, J. (1977). Sex differences in social participation. *Journal of Personality and Social Psychology, 35*, 742–754.

Zuroff, D. C., Stotland, S., Sweetman, E., Craig, J. A. & Koestner, R. (1995). Dependency, self-criticism, and social interactions. *British Journal of Clinical Psychology, 34*, 543–553.