

Probing Interactions in Fixed and Multilevel Regression: Inferential and Graphical Techniques

Daniel J. Bauer and Patrick J. Curran
University of North Carolina at Chapel

Many important research hypotheses concern conditional relations in which the effect of one predictor varies with the value of another. Such relations are commonly evaluated as multiplicative interactions and can be tested in both fixed- and random-effects regression. Often, these interactive effects must be further probed to fully explicate the nature of the conditional relation. The most common method for probing interactions is to test simple slopes at specific levels of the predictors. A more general method is the Johnson-Neyman (J-N) technique. This technique is not widely used, however, because it is currently limited to categorical by continuous interactions in fixed-effects regression and has yet to be extended to the broader class of random-effects regression models. The goal of our article is to generalize the J-N technique to allow for tests of a variety of interactions that arise in both fixed- and random-effects regression. We review existing methods for probing interactions, explicate the analytic expressions needed to expand these tests to a wider set of conditions, and demonstrate the advantages of the J-N technique relative to simple slopes with three empirical examples.

Substantive theories within psychology, education and many other disciplines in the social sciences are replete with instances in which the effect of one predictor on the outcome is hypothesized to vary as a function of a second predictor. We will refer to the first of these predictors as the *focal predictor* and the second as the *moderator*, although this distinction is strictly theoretical. Hypotheses concerning

This work was funded in part by fellowship DA06062 awarded to the first author and grant DA13148 awarded to the second author. We would like to thank R. J. Wirth and the members of the Carolina Structural Equations Modeling Group for their valuable input throughout this project.

Correspondence concerning this article should be addressed to Daniel J. Bauer, L.L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina, Chapel Hill, NC 27599–3270. E-mail: dbauer@email.unc.edu

moderated relationships can be evaluated in the standard fixed-effects regression model (e.g., Cohen, Cohen, West, & Aiken, 2003), or they can manifest in more complicated ways in the multilevel (or random effects) regression model (e.g., Raudenbush & Bryk, 2002). Regardless of the analytic approach, testing the unique contribution of the product term net the lower-order main effects provides an omnibus test of the interaction effect (Baron & Kenny, 1986; Cohen, 1978). More detailed information on the nature of the moderated relationship must then be obtained through the use of techniques for probing the interaction term (e.g., Aiken & West, 1991).

For the standard fixed-effects regression model, the most popular method for probing interactions is the “pick-a-point” approach (Rogosa, 1980). This approach involves plotting and testing the conditional effect of the focal predictor at designated levels of the moderating variable (e.g., high, medium, and low), where these conditional effect estimates are commonly referred to as “simple slopes” (Aiken & West, 1991; Jaccard & Turrisi, 2003). Although widely used in fixed-effects regression, the consideration of simple slopes in random-effects regression has been limited. While Hox (2002, pp. 58–63) recommended the plotting of simple slopes to aid in the interpretation of interactions in multilevel models, to our knowledge this approach has only been used in a small number of empirical applications and almost always as a descriptive rather than inferential device (e.g., Bryk & Raudenbush, 1987, p. 154; Hussong, 2003; Willett, Singer & Martin, 1998, p. 423; Yip & Fuligni, 2002).

While the pick-a-point approach is informative and easy to use, typically only a small number of specific values of the moderator are selected to evaluate the conditional effect of the focal predictor on the outcome. These values are often selected somewhat arbitrarily (Rogosa, 1980), and may even reside outside of the range of the sample data (e.g., one standard deviation above the mean of a skewed predictor). Interest may instead center on how the conditional effect of the focal predictor changes across the entire range of a continuous moderating variable, rather than just at a few values.

One such instance is when a grouping variable and continuous variable interact in an ANCOVA or linear regression model. In the presence of such an interaction, the group means cannot be compared by simply adjusting for differences on the continuous covariate. An alternative approach is to evaluate the group mean difference at each level of the covariate with the goal of determining when the group mean differences are and are not statistically different. The J-N technique was developed for this purpose (Johnson & Fay, 1950; Johnson & Neyman, 1936) and involves the use of two closely related procedures for evaluating conditional effects, the computation of regions of significance and confidence bands (Huitema, 1980). Whereas regions of significance define the levels of the covariate for which the group mean difference is significant, confidence bands convey the precision with which the group mean difference is estimated at each level of the covariate.

There are two key advantages of the J-N technique. First, regions of significance provide an inferential test for any possible simple slope of the focal predictor variable. Second, the confidence bands graphically depict the precision of estimation of the effect of the focal predictor over the full range of the moderator, providing a range of likely values for the conditional effect that narrows or widens depending on the selected level of the moderator. The point where the confidence bands are narrowest, also known as the *center of accuracy*, represents the value of the moderator at which we have the most confidence estimating the conditional effect of the focal predictor on the outcome.

Despite these distinct advantages of the J-N technique, the use of regions of significance and confidence bands for evaluating interactions remains extremely limited. One possible reason for this is that the J-N technique has not been generalized to settings other than the traditional dichotomous by continuous interaction in the ANCOVA model, even in such recent treatments as Cohen et al. (2003) and Jaccard and Turrisi (2003). For instance, to our knowledge there are no published applications in which this technique has been applied to evaluate continuous by continuous interactions in standard regression models. Further, aside from some preliminary developments by ourselves (Bauer, Curran, & Bollen, 2001; Curran, Bauer, & Willoughby, 2004, in press) and Miyazaki (2002), we are aware of no prior work that has extended these methods in a general way to the random-effects or multilevel regression model.¹ Our goal is to make these extensions here. We first present our expansion of the J-N technique for the more familiar fixed-effects regression model, then extend the use of this technique to the multilevel regression model involving random effects. Empirical examples are provided throughout to illustrate the use of these procedures.

PROBING INTERACTIONS IN FIXED-EFFECTS REGRESSION

A simple fixed-effects regression model involving two predictors and their interaction can be expressed in scalar notation as

$$y_i = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{1i} x_{2i} + \epsilon_i. \quad (1)$$

The coefficient γ_3 represents the interaction, defined as the effect of the $x_1 x_2$ product variable on the dependent variable y over and above the additive main effects of the predictors x_1 and x_2 (γ_1 and γ_2 , respectively). The residuals ϵ represent the portion of variability in y that is unexplained by the predictors.

¹Since the writing of this manuscript, related work on this topic has been published by Tate (2004).

For this model, we can define the *prediction equation* to be

$$\mu_{y|x_1, x_2} = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_1 x_2, \quad (2)$$

where the expected value of y conditioned upon specific values of x_1 and x_2 is denoted $\mu_{y|x_1, x_2}$. Finally, we can rearrange Equation 2 to highlight how the effect of one predictor varies as a function of the other. Treating x_1 as the focal predictor and x_2 as the moderator, we have

$$\mu_{y|x_1, x_2} = (\gamma_0 + \gamma_2 x_2) + (\gamma_1 + \gamma_3 x_2) x_1, \quad (3)$$

explicating that the y on x_1 regression line varies as a function of x_2 . We can then designate the *simple intercept* and *simple slope* of this conditional relationship as ω_0 and ω_1 , respectively:

$$\omega_0 = \gamma_0 + \gamma_2 x_2, \quad (4)$$

$$\omega_1 = \gamma_1 + \gamma_3 x_2. \quad (5)$$

One important fact that Equations 4 and 5 highlight is that, in the presence of a product interaction, the intercept and main effects of the regression model are scale-dependent. For instance, Equation 5 shows that the main effect of the focal predictor (γ_1) is the effect where the moderator is zero ($x_2 = 0$). Because zero is often outside the logical range of the moderator, it is recommended to mean center or standardize predictors involved in interactions so that the main effects represent the effects of each predictor at the mean level of the other (see Jaccard, Turrisi, & Wan, 1990, and Aiken & West, 1991, who also discuss computational advantages of mean centering predictors). While this aids in the interpretation of the main effect, we will typically also want to assess the effect of the focal predictor at levels of the moderator other than zero to more fully explore the nature of the interaction. The pick-a-point and J-N techniques provide alternative methods for accomplishing this goal. For both approaches, the key is to obtain an estimate and standard error for a parameter (e.g., ω_1) that is a linear composite of other parameters (e.g., γ_1 and γ_3). We now show how to compute these values for the simple slope of the focal predictor (ω_1), as it is usually this value that is of most interest, though similar procedures could also be applied to the simple intercept (ω_0).

Estimating Conditional Effects in Fixed-Effects Regression

In the general case, the fixed-effects regression model may be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{6}$$

where \mathbf{y} is an $n \times 1$ criterion vector, \mathbf{X} is the $n \times p$ design matrix for p -predictor variables (including a column vector of 1's to define the intercept), $\boldsymbol{\gamma}$ is a $p \times 1$ vector of fixed regression parameters relating each predictor to the criterion, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of random residuals. The residuals are assumed to be homoscedastic and normally distributed [i.e., $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$]. In the specific case of the two-predictor interaction model given in Equation 1, the matrix \mathbf{X} would simply consist of a column of 1's, a column of observed values for x_1 , a column of observed values for x_2 , and a column of values for the product of x_1 and x_2 .

From a frequentist perspective, the parameter estimates of the model may be viewed as random variables characterized by a joint sampling distribution. Given sufficiently large samples, this sampling distribution will obtain a multivariate normal form. Conditional effect estimates may then be viewed as weighted linear composites of these random normal variates (Aiken & West, 1991, pp. 25–26; Morrison, 1990, p. 83):

$$\hat{\omega} = a_1 \hat{\gamma}_0 + a_2 \hat{\gamma}_1 \dots + a_p \hat{\gamma}_{p-1} = \mathbf{a}' \hat{\boldsymbol{\gamma}}, \tag{7}$$

where \mathbf{a} is a $p \times 1$ column vector containing the a_1, a_2, \dots, a_p fixed weights used to form the composite. The variance of the conditional effect is estimated as

$$\widehat{VAR}(\hat{\omega}) = \widehat{VAR}(\mathbf{a}' \hat{\boldsymbol{\gamma}}) = \mathbf{a}' \widehat{ACOV}(\hat{\boldsymbol{\gamma}}) \mathbf{a}, \tag{8}$$

where $\widehat{ACOV}(\hat{\boldsymbol{\gamma}})$ is the sample estimate of the asymptotic covariance matrix of the regression coefficient estimates given by the usual formula

$$\begin{aligned} \widehat{ACOV}(\hat{\boldsymbol{\gamma}}) &= [\mathbf{X}'(\hat{\sigma}^2 \mathbf{I}_n)^{-1} \mathbf{X}]^{-1} \\ &= \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{9}$$

The standard error of the estimate is then simply the square root of the variance estimate obtained from Equation 8.

For our simple two-predictor example the vector of regression coefficient estimates is

$$\hat{\boldsymbol{\gamma}}' = (\hat{\gamma}_0 \quad \hat{\gamma}_1 \quad \hat{\gamma}_2 \quad \hat{\gamma}_3). \quad (10)$$

Thus, to obtain a point estimate and standard error for ω_1 as defined in Equation 5, we would set

$$\mathbf{a}' = (0 \quad 1 \quad 0 \quad x_2), \quad (11)$$

and solve Equations 7 and 8 to obtain the point estimate and variance of the simple slope where

$$\hat{\omega}_1 = \hat{\gamma}_1 + \hat{\gamma}_3 x_2, \quad (12)$$

$$\widehat{VAR}(\hat{\omega}_1) = \widehat{VAR}(\hat{\gamma}_1) + 2x_2 \widehat{COV}(\hat{\gamma}_1, \hat{\gamma}_3) + x_2^2 \widehat{VAR}(\hat{\gamma}_3). \quad (13)$$

Equations 12 and 13 represent the two fundamental pieces of information that will be manipulated to probe the interaction effect.

We next briefly review the testing of simple slopes using the pick-a-point approach, and then proceed to detail how the J-N technique can be applied to more fully evaluate conditional effects resulting from the interaction of categorical or continuous predictors.

The Pick-a-Point Approach: Tests of Simple Slopes

A traditional test of simple slopes using the pick-a-point technique consists of computing conditional effect estimates $\hat{\omega}$ and their variances $\widehat{VAR}(\hat{\omega})$ using Equations 7 and 8 for several fixed values of the moderating variable (e.g., 0 and 1 if the moderator is binary, or high, medium, and low values if it is continuous). A test of a simple slope is then obtained by forming the critical ratio of the conditional effect estimate to its standard error at a specified level of the moderator:

$$t = \frac{\hat{\omega}}{[\widehat{VAR}(\hat{\omega})]^{1/2}}. \quad (14)$$

Cohen et al. (2003) and Jaccard and Turrissi (2003) have also suggested constructing confidence intervals for simple slope estimates rather than relying exclusively on null hypothesis tests, a point that we will return to and reinforce shortly. In addition, a graphical depiction of the interaction may also be obtained by computing

both simple intercepts and simple slopes for each of several levels of the moderator and then plotting the simple regression lines given by these estimates.

While this approach has been an extremely useful method for probing interactions in many areas of applied research, it is typically limited to the evaluation of conditional effects at only a small number of values of the moderator. There is not a systematic way to examine the magnitude of the conditional effects, nor their precision, across the entire range of the moderator, information that is provided by the more general J-N technique. For this reason, we now turn to a further explication of the J-N technique in both fixed-effects and random-effects regression.

The J-N Technique

We distinguish between two related aspects of the J-N technique: The computation of regions of significance and the plotting of confidence bands for the conditional effect. To our knowledge, this presentation is the first to expand the use of the J-N technique beyond the traditional ANCOVA framework. In our view, this expansion has been impeded by the traditional formulation of the J-N technique from separate within-group linear regressions of the outcome on the covariate (see Aiken & West, 1991, pp. 134–136; Jaccard & Turrisi, 2003, pp. 81–82). Extension of the technique to continuous by continuous interactions is then nonobvious, given the lack of discretely defined groups. However, as has been shown elsewhere (Hunka, 1995; Rogosa, 1980, 1981) the traditional J-N technique can be embedded in a general linear model involving an interaction between the grouping variable and continuous covariate. It then becomes a relatively simple matter to consider use of the J-N technique for interactions between continuous variables as well. We thus motivate the J-N technique from the standpoint of the general fixed-effects regression model. This presentation has the added benefit of explicating the analytical relation between the J-N technique and the pick-a-point approach which in turn facilitates our ultimate goal of extending the J-N technique to a broad class of random-effects regression models.

Regions of significance. The computation of regions of significance is an alternative to the pick-a-point approach that indicates *over what range of the moderator* the effect of the focal predictor is significantly positive, nonsignificant, or significantly negative. To compute regions of significance, we reverse the unknown quantity in the critical ratio in Equation 14. Specifically, rather than solve for the t value that corresponds to the simple slope at the designated level of the moderator, we select a critical value t_{crit} (i.e., ± 1.96 for large samples) and solve for the values of the moderator that return that critical value:

$$\pm t_{crit} = \frac{\hat{\omega}}{[\widehat{VAR}(\hat{\omega})]^{1/2}}. \tag{15}$$

Manipulation of this equation yields

$$t_{crit}^2 \widehat{VAR}(\hat{\omega}) - \hat{\omega}^2 = 0. \quad (16)$$

For interactions involving two predictors, the two roots of the moderator that satisfy this equality can be solved by the quadratic formula. These roots demarcate the boundaries of the regions of significance, indicating the points on the scale of the moderator at which the effect of the focal predictor passes from significance to nonsignificance at the selected alpha-level.

For example, let us again consider the conditional effect of x_1 in the simple two predictor model in Equation 1. Substituting in the expressions in Equations 12 and 13 into Equation 16 and collecting terms, we have

$$ax_2^2 + bx_2 + c = 0, \quad (17)$$

where

$$a = t_{crit}^2 \widehat{VAR}(\hat{\gamma}_3) - \hat{\gamma}_3^2, \quad (18)$$

$$b = 2[t_{crit}^2 \widehat{COV}(\hat{\gamma}_1, \hat{\gamma}_3) - \hat{\gamma}_1 \hat{\gamma}_3], \quad (19)$$

$$c = t_{crit}^2 \widehat{VAR}(\hat{\gamma}_1) - \hat{\gamma}_1^2. \quad (20)$$

The values of x_2 satisfying this equality can then be obtained via the quadratic formula

$$x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (21)$$

The two values (or roots) of x_2 that are returned by this formula demarcate the boundaries of the regions of significance.² Note that this holds regardless of the distribution of the moderator. The moderator may thus be categorical or continu-

²Sometimes the conditional effect of the focal predictor will be nonsignificant at any given level of the moderator despite a significant test of the product interaction term. Mathematically, this will result in there being no real roots to Equation 21 (i.e., the roots will be imaginary numbers). The opposite can also occur, where regions of significance and nonsignificance can be identified despite the lack of a significant interaction term. Rogosa (1980, 1981) provides a useful discussion of these scenarios.

ous, making the traditional application of the J-N technique in the ANCOVA model a special case of Equation 16.

Confidence bands. One common feature of tests of simple slopes and regions of significance is that both procedures are based on traditional null hypothesis testing. However, the usefulness of this approach to statistical inference has been questioned repeatedly over the decades, culminating in the APA task force report emphasizing that confidence intervals are much more informative than null hypothesis tests (Wilkinson & the Task Force on Statistical Inference, 1999). Similarly, as we now show, confidence bands provide more information than null hypothesis tests of simple slopes and regions of significance.

The standard formula for a confidence interval (CI) is

$$CI = \hat{\theta} \pm t_{crit} [\widehat{VAR}(\hat{\theta})]^{1/2}, \tag{22}$$

where $\hat{\theta}$ represents a given point estimate. In most cases we are interested in a single effect estimate and so simply compute the confidence interval for this estimate. However, in the case of conditional effects, both the effect estimate and its standard error vary as a function of the moderating variable. As such, we cannot plot just one confidence interval; instead we must plot the confidence interval over the full range of the moderating variable, and these are known as confidence bands (Rogosa, 1980, 1981). The general formula for the confidence bands (CB) is then

$$CB_{\hat{\omega}} = \hat{\omega} \pm t_{crit} [\widehat{VAR}(\hat{\omega})]^{1/2}. \tag{23}$$

To continue with our simple two predictor model, if we were interested in computing confidence bands for the effect of the focal predictor x_1 as a function of the moderator x_2 , we would substitute in the specific conditional effect estimate of x_1 and its variance as given in Equations 12 and 13:

$$CB_{\hat{\omega}_1} = (\hat{\gamma}_1 + \hat{\gamma}_3 x_2) \pm t_{crit} [\widehat{VAR}(\hat{\gamma}_1) + 2x_2 \widehat{COV}(\hat{\gamma}_1, \hat{\gamma}_3) + x_2^2 \widehat{VAR}(\hat{\gamma}_3)]^{1/2}. \tag{24}$$

As is the case with standard confidence intervals, these confidence bands convey the same information as null hypothesis tests of simple slopes and/or regions of significance. Specifically, the points where the confidence bands cross zero are the boundaries of the regions of significance (also indicating which simple slopes are significant and which are not).

Additionally, the confidence bands graphically convey our certainty in the conditional effect estimates and how that certainty changes as we progress across the range of the moderating variable. The confidence bands are narrowest at the *center of accuracy*. This point can be obtained by differentiating the variance function for the conditional effect estimate with respect to the moderating variable and setting the resulting derivative equal to zero

$$\frac{d}{dx_m} \widehat{VAR}(\hat{\omega}) = 0. \quad (25)$$

Solving for values of the moderator, here denoted x_m , that satisfy this equality will produce candidate points for the center of accuracy. For our simple example model, manipulation of Equation 25 shows that the center of accuracy is the value of x_2 given by the formula

$$x_2 = \frac{-\widehat{COV}(\hat{\gamma}_1, \hat{\gamma}_3)}{\widehat{VAR}(\hat{\gamma}_3)}. \quad (26)$$

More complicated variance functions may have more than one candidate point for the center of accuracy, requiring that these points be compared to determine the one yielding the smallest variance. Typically, the center of accuracy will be in the middle of the scale of the moderator, with decreasing precision evident at the ends of the scale.

Multiple Testing: Nonsimultaneous Versus Simultaneous Inference

The regions and bands explicated above, as well as the pick-a-point procedure described before them, are all predicated on the use of an alpha-level that is only valid for a single test. The inferences they afford must thus be regarded as nonsimultaneous. As such, the Type I error rates of these techniques are only fully accurate for a single selected value of the moderator and the error rate will otherwise accumulate as additional values of the moderator are considered. If one wished to simultaneously evaluate the effect of the focal predictor at a small number of values of the moderator (e.g., high, medium, and low, as in tests of simple slopes), a family-wise alpha-level could be obtained by using, for instance, the Bonferroni approach (Neter, Kutner, Nachtsheim, & Wasserman, 1996, pp. 157–158). This approach could similarly be applied to the regions and confidence bands of the J-N technique so that inferences would be valid for a specific number of values of the moderator. If, with the J-N technique, valid in-

ferences are sought for all values of the moderator simultaneously, then one can instead use the critical values derived by Potthoff (1964) to compute simultaneous regions of significance and confidence bands. In the simple case of the model considered above, the critical t value in Equations 15 through 24 would be replaced by $[2F_{crit}(2, N - p)]^{1/2}$ (see also Rogosa, 1980, 1981, and Serber, 1977). If the number of comparisons to be made is not small, this approach will generally be more powerful than using the Bonferroni approach and has the further advantage of yielding an accurate Type I error rate for all possible comparisons (Neter et al., 1996, pp. 157–158).

Despite these advantages, the simultaneous critical value of Potthoff (1964) will necessarily yield narrower regions of significance and broader confidence bands than the nonsimultaneous critical value, sometimes so much so that the results may be of little practical use, as noted by Potthoff himself. As a possible remedy to the problem, Potthoff suggested using a higher alpha level with the simultaneous approach. Others have suggested alternative forms for the confidence bands that are valid only over specific (observed) intervals of the moderating variable, rather than over the entire real line (Aitken, 1973; Gafarian, 1964). Unless otherwise noted, here we focus specifically on the nonsimultaneous version of the J-N technique for the principal reasons that this facilitates an understanding of the relation between the pick-a-point and J-N procedures, no consensus has yet emerged on the best method for constructing simultaneous regions and confidence bands, and the simultaneous approach is often impractical. Mindful of the error rate accumulation problem of the nonsimultaneous approach, however, we view the resulting regions and bands as primarily of heuristic value.

Empirical Example: The Relation Between Child Math Ability, Delinquency, and Hyperactivity

To demonstrate the application of the J-N technique to probe a continuous by continuous interaction in a fixed-effects regression model, we considered a sample of $N = 956$ children drawn from the 1990 assessment of the Children of the National Longitudinal Survey of Youth (NLSY). Children ranged in age from 59 to 156 months (mean age of 100 months), ranged in grade from kindergarten to sixth grade (third grade was the median), 52% were female, and 55% of mothers reported their child to be of minority status (self identified to be non-Caucasian). The motivating substantive question related to the prediction of math ability as a function of the joint influence of antisocial and hyperactive behavior. It was hypothesized that there would be a negative relation between child antisocial behavior and math ability, and that this effect would be potentiated by the additional presence of child hyperactive behavior.

The regression model used to evaluate this hypothesis was identical to Equation 1, with the exception that age, grade, sex, and minority status were entered

TABLE 1
 Results From Fixed-Effects Regression of Math Ability on Antisocial Behavior, Hyperactive Behavior, and the Multiplicative Interaction Between Antisocial and Hyperactive Behavior

<i>Fixed Effects</i>	<i>PE</i>	<i>SE</i>	<i>p</i>
Intercept ($\hat{\gamma}_0$)	38.07	.322	< .0001
Antisocial ($\hat{\gamma}_1$)	.0373	.2681	.8893
Hyperactive ($\hat{\gamma}_2$)	-.799	.2148	.0002
Antisocial by Hyperactive ($\hat{\gamma}_3$)	-.397	.1429	.0055
Age	.347	.041	< .0001
Gender	-2.09	.592	.0004
Grade	3.40	.519	< .0001
Minority status	-3.70	.583	< .0001

<i>Asymptotic Variance/Covariance Matrix of Fixed Effects</i>				
	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
$\hat{\gamma}_0$.1039			
$\hat{\gamma}_1$.0129	.0719		
$\hat{\gamma}_2$	-.0003	-.0268	.0461	
$\hat{\gamma}_3$	-.0214	-.0124	.000	.0204

Note. All models were fit using SAS PROC REG using the sample of $N = 956$ children drawn from the NLSY described in the text. The parameter estimates and standard errors for the main effects of age, gender, grade, and minority status are shown above, but these have no direct relevance with respect to testing and probing the interaction term and are thus not included in the reported ACOV matrix. PE = parameter estimate; SE = standard error.

as additional covariates. Here, y is mathematics ability, measured as the proportion of 84 items correctly endorsed on the math subsection of the Peabody Individual Achievement Test ($\bar{y} = 37.65$, $sd = 15.77$, $range = 5.95 - 82.14$), x_1 is a continuous measure of antisocial behavior as reported by the child's mother ($\bar{x}_1 = 1.34$, $sd_{x_1} = 1.34$, $range = 0 - 6$), x_2 is a continuous measure of hyperactive behavior as reported by the child's mother ($\bar{x}_2 = 1.94$, $sd_{x_2} = 1.54$, $range = 0 - 5$), and larger values of these variables reflect higher levels of child antisocial and hyperactive behavior. We mean-centered all predictors (including the covariates) prior to fitting the model using ordinary least squares estimation.³

Together, the predictors accounted for 69% of individual differences in math ability (see Table 1), $F(7, 948) = 299.28$, $p < .0001$, $R^2_{adj} = .69$. To further probe the significant interaction, we computed the simple intercepts and simple slopes for the regression of math ability on antisocial behavior at the mean of hyperactivity

³This centering included the dichotomous measures of minority status and female. This allowed for the calculation of simple intercepts assessed at the mean of all predictors, and not assessed within the groups coded as zero on these predictors.

and at one standard deviation above and below the mean, producing the plots in Figure 1. Only the simple slope of antisocial behavior at high hyperactivity was statistically significant ($p = .045$); the relation of antisocial behavior to mathematics ability was not significantly different from zero at medium and low levels of hyperactivity. Although these simple regressions substantially aid in the interpretation of this interactive effect, the J-N technique can provide additional useful information.

Applying Equations 16 and 23 resulted in 95% (nonsimultaneous) regions of significance defined by a lower bound of -2.32 and an upper bound of 1.49 . As shown in Figure 2, these regions imply that the regression of math ability on antisocial behavior is significant and positive at values of hyperactivity less than -2.32 , not significantly different from zero at values of hyperactivity between -2.32 and 1.49 , and significant and negative at values of hyperactivity greater than 1.49 . Given that the minimum and maximum values of (mean centered) hyperactivity were -1.94 and 3.06 , respectively, the upper region fell within the observed range of hyperactivity whereas the lower region did not (hence it will not be interpreted further). Indeed, 167 of the 956 children (17.5%) had hyperactivity scores greater than the upper bound of the region. The 95% confidence bands that correspond to these regions are also presented in Figure 2. The center of accuracy, the narrowest cross-section of the confidence bands, was determined to be $.61$, a point

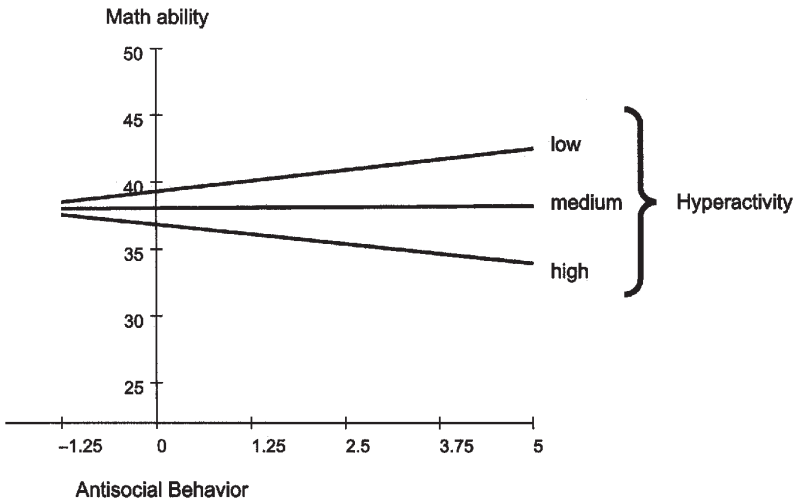


FIGURE 1 The pick-a-point simple slopes of the regression of math ability on antisocial behavior at high, medium, and low levels of hyperactive behavior. *Note.* High, medium, and low values of hyperactivity are defined as plus and minus 1 *sd* about the mean (1.54, 0, -1.54). The slope of the simple regression at high hyperactivity significantly differs from zero, but the simple slope at medium and low hyperactivity does not.

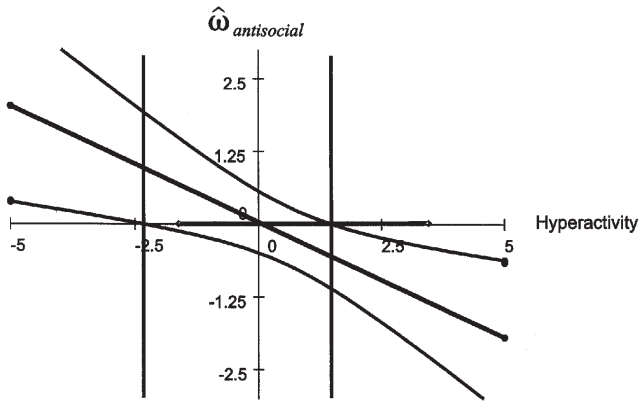


FIGURE 2 J-N regions of significance and confidence bands for the conditional relation between math ability and antisocial behavior as a function of hyperactivity. *Note.* Dashed vertical lines reflect regions of significance (-2.32, 1.49) and the dark horizontal line with diamonds indicates the range of child hyperactivity observed in the sample data (-1.94, 3.06). The intersection of values of hyperactivity equal to -1.54, 0, and 1.54 with effect of antisocial behavior correspond to the three simple slopes presented in Figure 1.

where the effect of antisocial behavior is nonsignificant. For contrast, the simultaneous regions of significance were also calculated (not shown in Figure 2), with obtained roots of -5.52 and 2.18. The lower root is well outside the observed range of the data, and only 7% of the children's hyperactivity scores were observed above the upper root. The smaller region of significance with the simultaneous approach is a natural consequence of trade-off between simultaneous inference and power, but in this case the region includes such a small portion of the sample that it is of relatively little practical use.

In summary, the J-N technique provides much additional information relative to the pick-a-point approach. Indeed, the testing of specific simple slopes is also provided by the J-N technique (e.g., the simple slopes of the high, medium and low regression lines plotted in Figure 1, and their confidence intervals, simply reflect three specific points on the abscissa in Figure 2). In addition, these regions convey that the effect of antisocial behavior on mathematics ability is significant at any given level of hyperactivity higher than 1.49 units above the mean and that we have the most confidence estimating this effect for hyperactivity levels slightly above the mean. All of these results, and the corresponding plots, can be calculated using a suite of freely available javascript programs accessible at <http://www.quantpsy.org> that implement the formulas presented in this paper. Table 1 provides all the necessary information for the interested reader to replicate our results using this program. We now turn to the extension of the J-N technique to a broad class of random-effects regression models.

PROBING INTERACTIONS
IN MULTILEVEL REGRESSION

Unlike the standard fixed-effects regression models, interactions can be manifested in a variety of ways in multilevel models. They may occur within a given level of the model (e.g., an interaction between two lower-level or two upper-level variables) or between levels of the model (e.g., an interaction between a lower-level variable and an upper-level variable). Moreover, some interactions will involve predictors whose main effects are random, or even the interaction may be a random effect, whereas in other cases these effects will be fixed. The most common of these scenarios is the cross-level interaction with a random main effect for the lowest level predictor. We use this case to show how the J-N technique can be extended and applied to multilevel models, but of course these procedures directly apply to the other types of interactions that can arise in multilevel models as well.

A cross-level interaction arises when the random coefficient of a lower level predictor is itself predicted by an upper level predictor. In the simplest case, the level-1 model can be written as

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \epsilon_{ij}, \tag{27}$$

where y_{ij} is the outcome variable for individual i in group j , x_{1ij} is the single predictor for individual i in group j , β_{0j} and β_{1j} are the intercept and slope of the regression of y on x_1 within group j , respectively, and ϵ_{ij} is the random residual for individual i within group j . Given that the intercept and slope coefficients vary randomly over groups, these can be conceived as random variables to be predicted by one or more group-level covariates. In the case of a single group-level covariate, the level-2 model can be written as

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + u_{0j}, \tag{28}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{1j} + u_{1j}, \tag{29}$$

where w_{1j} represents the single predictor for group j , γ_{00} and γ_{10} are the fixed intercepts of the regression of β_{0j} and β_{1j} on w_{1j} , and γ_{01} and γ_{11} represent the fixed slopes of the regression of β_{0j} and β_{1j} on w_{1j} , and u_{0j} and u_{1j} reflect the residual variability in the level-1 intercepts and slopes net the prediction by w_{1j} .

Although there is only a single main effect in the level-1 equation and a single main effect in each level-2 equation, the cross-level interaction effect is apparent in

the reduced-form equation, obtained by substituting Equations 28 and 29 into Equation 27 and grouping fixed and random components:

$$y_{ij} = (\gamma_{00} + \gamma_{01}w_{1j} + \gamma_{10}x_{1ij} + \gamma_{11}x_{1ij}w_{1j}) + (u_{0j} + u_{1j}x_{1ij} + \epsilon_{ij}). \quad (30)$$

This equation illustrates that the main effect regression of the random slopes on w_1 (γ_{11}) is expressed as a cross-level interaction between w_1 and x_1 in the reduced form equation (i.e., $\gamma_{11}w_{1j}x_{1ij}$). Given the formulation of the level-1 and level-2 models, we would traditionally view w_1 as the moderator of the effect of x_1 . However, as always, the conditional effects in the model are symmetrical, and in the reduced-form model we can see that there is no restriction on which variable is to be considered the focal predictor or moderator.

To consider the conditional nature of the effects of each predictor in greater detail, we first write the prediction equation for the model. Taking the expectation of Equation 30 over both individuals and groups yields the prediction equation

$$\mu_{y|x_1, w_1} = \gamma_{00} + \gamma_{10}x_1 + \gamma_{01}w_1 + \gamma_{11}x_1w_1. \quad (31)$$

The simple slope for the regression of y on x_1 as a function of w_1 is then

$$\omega_{x_1} = \gamma_{10} + \gamma_{11}w_1. \quad (32)$$

Similarly, given the symmetry of the interaction, the simple slope of the regression of y on w_1 can be written as a function of x_1 , or

$$\omega_{w_1} = \gamma_{01} + \gamma_{11}x_1. \quad (33)$$

As for the standard regression model, the challenge is then again to obtain an estimate and standard error for a linear composite. Given that this linear composite consists entirely of fixed effects, we will see that this task is accomplished in much the same way for the multilevel regression model as it was for the standard regression model.

Estimating Conditional Effects in Multilevel Models

More formally, the reduced form of the linear multilevel model can be expressed as

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\gamma} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j, \quad (34)$$

where \mathbf{y}_j is the $n_j \times 1$ response vector for group $j = 1, 2, \dots, J$, \mathbf{X}_j is the $n_j \times p$ design matrix for the $p \times 1$ vector of fixed effects $\boldsymbol{\gamma}$ (where p includes a column vector of

1's for the intercept), \mathbf{Z}_j is the $n_j \times q$ design matrix for the $q \times 1$ vector of random effects \mathbf{u}_j , and $\boldsymbol{\epsilon}_j$ is the $n_j \times 1$ vector of residuals (Laird & Ware, 1982). Importantly, it is assumed that the random effects and residuals are independent of one another and are multivariate normally distributed as

$$\mathbf{u}_j \sim N(\mathbf{0}, \mathbf{T}), \tag{35}$$

$$\boldsymbol{\epsilon}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_j}). \tag{36}$$

Although not requisite, the form of the covariance matrix of the random effects \mathbf{T} is typically unrestricted and the residuals are often constrained to be homoscedastic and independent (i.e., $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_j} = \sigma^2 \mathbf{I}_{n_j}$). Without loss of generality we follow these conventions here, although all of our developments apply to any structure of \mathbf{T} and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_j}$ (assuming these are properly identified).

As we demonstrated earlier, once we move to the prediction equation by taking expectations over both individuals and groups, we are left with conditional effects that are linear composites of fixed coefficients only. Thus, the estimated conditional effect of x at a given level of w (or vice versa) can be computed from Equation 7 just as in the case of the standard regression model. Similarly, application of Equation 8 will yield the variance of this estimate. Where these computations differ from the standard fixed-effects case is that the asymptotic covariance matrix of the fixed-effects estimates no longer has the simple form of Equation 9, and instead is estimated as

$$\widehat{ACOV}(\hat{\boldsymbol{\gamma}}) = \left[\sum_{j=1}^J \mathbf{X}'_j \hat{\mathbf{V}}_j^{-1} \mathbf{X}_j \right]^{-1}, \tag{37}$$

where $\hat{\mathbf{V}}_j$ is the model-implied covariance matrix for \mathbf{y}_j , estimated as

$$\hat{\mathbf{V}}_j = \mathbf{Z}_j \hat{\mathbf{T}} \mathbf{Z}'_j + \hat{\sigma}^2 \mathbf{I}_{n_j}. \tag{38}$$

Equation 37 would simplify to Equation 9 if $\hat{\mathbf{T}}$ (or equivalently \mathbf{Z}) were a null matrix, that is, in the absence of random effects.

For example, consider the cross-level interaction in the multilevel regression model in Equation 30. Arranging the vector of fixed effect parameter estimates as

$$\hat{\boldsymbol{\gamma}}' = (\hat{\gamma}_{00} \quad \hat{\gamma}_{01} \quad \hat{\gamma}_{10} \quad \hat{\gamma}_{11}), \tag{39}$$

the estimate for the conditional fixed effect given in Equation 32 can be obtained from Equation 7 by defining

$$\mathbf{a}' = (0 \quad 0 \quad 1 \quad w_1). \tag{40}$$

This results in the linear composite

$$\hat{\omega}_{x_1} = \hat{\gamma}_{10} + \hat{\gamma}_{11}w_1. \tag{41}$$

Application of Equation 8 provides the estimated variance of this composite

$$\widehat{VAR}(\hat{\omega}_{x_1}) = \widehat{VAR}(\hat{\gamma}_{10}) + 2w_1\widehat{COV}(\hat{\gamma}_{10}, \hat{\gamma}_{11}) + w_1^2\widehat{VAR}(\hat{\gamma}_{11}). \tag{42}$$

Given this information, extension of the J-N technique to multilevel models is straightforward, as we now demonstrate.

Evaluating Interactions in Multilevel Models

Aside from requiring a more complex asymptotic covariance matrix to compute the variance of conditional effect estimates, tests of simple slopes, computation of regions of significance, and the plotting of confidence bands are all accomplished in the same way for multilevel models as for standard regression models, that is, via Equations 14, 16, and 23. For instance, to identify the regions of significance for the simple cross-level interaction in Equation 30, we would simply substitute the expressions in Equations 41 and 42 into Equation 16. Collecting terms, we have

$$aw_1^2 + bw_1 + c = 0, \tag{43}$$

where

$$a = t_{crit}^2 \widehat{VAR}(\hat{\gamma}_{11}) - \hat{\gamma}_{11}^2, \tag{44}$$

$$b = 2[t_{crit}^2 \widehat{COV}(\hat{\gamma}_{10}, \hat{\gamma}_{11}) - \hat{\gamma}_{10}\hat{\gamma}_{11}], \tag{45}$$

$$c = t_{crit}^2 \widehat{VAR}(\hat{\gamma}_{10}) - \hat{\gamma}_{10}^2. \tag{46}$$

As before, the two values of w_1 that satisfy this equality represent the boundaries of the regions of significance and can be obtained via the quadratic formula. Note that this again holds regardless of whether the moderator is categorical or continuous.

Similarly, substituting the expressions in Equations 41 and 42 into Equation 23, we see that the confidence bands are defined by the function

$$CB_{\hat{\omega}_{x_1}} = (\hat{\gamma}_{10} + \hat{\gamma}_{11}w_1) \pm t_{crit} \left[\widehat{VAR}(\hat{\gamma}_{10}) + 2w_1 \widehat{COV}(\hat{\gamma}_{10}, \hat{\gamma}_{11}) + w_1^2 \widehat{VAR}(\hat{\gamma}_{11}) \right]^{1/2}. \tag{47}$$

The center of accuracy can be determined by solving Equation 25, in this case resulting in the formula

$$w_1 = \frac{-\widehat{COV}(\hat{\gamma}_{10}, \hat{\gamma}_{11})}{\widehat{VAR}(\hat{\gamma}_{11})}. \tag{48}$$

The value of w_1 satisfying this equality is the value at which the conditional effect of x_1 is estimated with greatest certainty.

Despite the apparent ease with which the J-N technique generalizes to multi-level models, it is important to note one key complication. Under standard assumptions, the test statistics obtained from a standard regression model are exactly t -distributed with $n - p$ degrees of freedom. Unfortunately, the same is not typically true for a multilevel model; even under standard assumptions, the test statistics for the fixed effects are typically only approximately t -distributed (Kacker & Harville, 1984; Schaalje, McBride, & Fellingham, 2002). By implication, while exact tests of simple slopes, boundaries to the regions of significance, and width of the confidence bands can be obtained in the absence of random effects, when random effects are present the same computations provide only approximately valid inferences. As such, appropriate caution should be exercised in the interpretation of the results of these procedures in multilevel models. In particular, one should not place too much emphasis on the specific values obtained for the boundaries to the regions of significance, as these boundaries would likely change if another method for approximating the test distribution was selected (e.g., an alternative method for determining degrees of freedom). For cross-level interactions, we expect that the results will be most sensitive to this choice when there is a small sample size at the upper level of the model.

Similarly, while the distinction between nonsimultaneous and simultaneous versions of the J-N technique also arises in the multilevel model, the test statistic is not as easily derived as in the fixed-effects regression model. Specifically, Miyazaki (2002) argued that a critical F would be inappropriate given the more complex error structure of the multilevel model and has instead advocated the use

of a Wald test that is chi-square distributed when the number of groups J is large. Given that many applications of multilevel models involve a relatively small number of groups, the asymptotic nature of this test may provide a serious impediment to the simultaneous approach as outlined by Miyazaki. For this reason, and those adumbrated previously, we continue to use the nonsimultaneous version of the J-N technique here. However, whether one prefers a nonsimultaneous or simultaneous approach, it is fair to say that more study is needed to determine the optimal method for testing conditional effects in small samples so as to obtain accurate and stable estimates of the regions of significance and confidence bands.⁴

In summary, the principal difficulties associated with applying the J-N technique to multilevel models are no different than those that impact any multilevel modeling analysis, regardless of the presence of within-level or cross-level interactive effects. Specifically, the asymptotic covariance matrix of the fixed effects must take account of the additional variance components in the model and the test distribution for the fixed effects estimates is only approximately known so tests are typically inexact. Neither of these issues detracts from our ability to apply the J-N technique to evaluate conditional effects, and much interpretive information is to be gained from their use, as we now demonstrate with an empirical application.

EMPIRICAL EXAMPLE

To demonstrate the use of J-N regions and bands in practice, we considered data from the High School and Beyond (HSB) study that was described in detail by Raudenbush and Bryk (2002) and Singer (1998). These data consist of a total sample of 7,185 students who are nested within 160 schools. Between 14 to 67 students were assessed from each school, with a median of number of 47 students assessed. The outcome of interest is a child-level measure of math achievement ($\bar{y} = 12.75$, $sd = 6.89$). The first predictor was a continuous measure of child socioeconomic status (SES) which was group mean centered ($sd = .66$). The second predictor was a dichotomous measure of school sector in which a value of 0 reflected a public school and a value of 1 reflected a private school (49% of schools were private). The final predictor was a continuous measure of disciplinary climate of the school in which higher values reflected greater disciplinary problems (see Bryk & Thum, 1989, for further details). School discipline was grand mean centered ($sd = .94$).

We considered two separate models to demonstrate both a dichotomous by continuous cross-level interaction and a continuous by continuous cross-level interac-

⁴As one anonymous reviewer noted, a likelihood ratio test constitutes a plausible alternative to a t test of the conditional effect estimate, both for multilevel and standard regression models. A potentially fruitful direction for future research on this topic may be to generate methods for computing regions of significance and confidence bands based on likelihood ratio tests. Because likelihood ratio tests are only asymptotically chi-square distributed, we can anticipate that small sample sizes may also be problematic for this approach.

tion. Both models considered child SES as the sole level-1 predictor, but the first included sector as the sole level-2 predictor while the second included school discipline as the sole level-2 predictor. These models are thus of the form given in Equations 27 through 30. Although basic, these examples allow us to highlight the probing of two common types of cross-level interactions, and these methods generalize to any number of more complex (and more realistic) conditions. To be consistent with Singer’s (1998) analysis of the same data, both models were estimated in SAS Proc Mixed with restricted ML estimation and using the “between-within” method for computing degrees of freedom for tests of the fixed effects estimates (SAS Institute, 1999). However, the choice of method for computing degrees of freedom is relatively unimportant here, given the large number of level-2 units.

Child SES and school sector. The first multilevel model included the continuous measure of child SES as the sole level-1 predictor and the dichotomous measure of school sector as the sole level-2 predictor. Random effects were estimated for the level-1 intercept and slope, and both of these effects were regressed on school sector. Detailed results are presented in Table 2. All of the fixed effects were signifi-

TABLE 2
Results From HSB Multilevel Models With Categorical
by Continuous Interaction

<i>Model 1: Child SES and Sector</i>				
	<i>PE</i>	<i>SE</i>	<i>df</i>	<i>p</i>
Random effects				
Residual ($\hat{\sigma}^2$)	36.71	.63		< .0001
Intercept ($\hat{\tau}_{00}$)	6.73	.86		< .0001
Slope ($\hat{\tau}_{11}$)	.27	.23		.1228
Covariance ($\hat{\tau}_{01}$)	1.05	.34		.0021
Fixed effects				
Intercept ($\hat{\gamma}_{00}$)	11.39	.29	158	< .0001
Child SES ($\hat{\gamma}_{10}$)	2.80	.16	7023	< .0001
Sector ($\hat{\gamma}_{01}$)	2.81	.44	158	< .0001
Child SES by Sector ($\hat{\gamma}_{11}$)	-1.34	.23	7023	< .0001
-2LL			46638.6	
<i>Asymptotic Variance/Covariance Matrix of Fixed Effects</i>				
	$\hat{\gamma}_{00}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{01}$	$\hat{\gamma}_{11}$
$\hat{\gamma}_{00}$.086			
$\hat{\gamma}_{10}$.012	.024		
$\hat{\gamma}_{01}$	-.086	-.012	.193	
$\hat{\gamma}_{11}$	-.012	-.024	.027	.055

Note. All models were fit using SAS PROC MIXED using the High School and Beyond data described in the text. PE = parameter estimate; SE = standard error.

cant, most notably the cross-level interaction between child SES and school sector reflecting that the magnitude of the relation between child SES and math achievement varied as a function of school sector.

To probe this effect, we computed the simple slopes of math achievement on child SES within each level of sector. Results indicated that there was a significantly positive relation between math achievement and child SES for both sectors, but this relation was significantly stronger for children enrolled in public schools. Further, children enrolled in private schools reported significantly higher math scores at lower levels of SES, but the magnitude of this effect diminished with increasing SES (see Figure 3). However, the *specific point* on child SES at which the difference between private and public schools becomes nonsignificant is not known using the pick-a-point approach. To identify this point, we applied the J-N technique to calculate the regions of significance and associated confidence bands (see Figure 4).

The boundaries to the regions of significance indicated that math achievement scores were significantly higher for children enrolled in private schools when child SES was less than 1.23, the difference between sectors was nonsignificant between SES values of 1.23 and 3.65, and children in public schools outperformed those in private schools at values of SES greater than 3.65. Given that the observed values on SES ranged from -3.65 to 2.86 (indicated in the figure by the darkened portion of the abscissa), these regions indicated that in the HSB sample, children in private schools reported significantly higher math achievement scores at any given value of SES up to 1.89 standard deviations above the mean, but this effect was not sig-

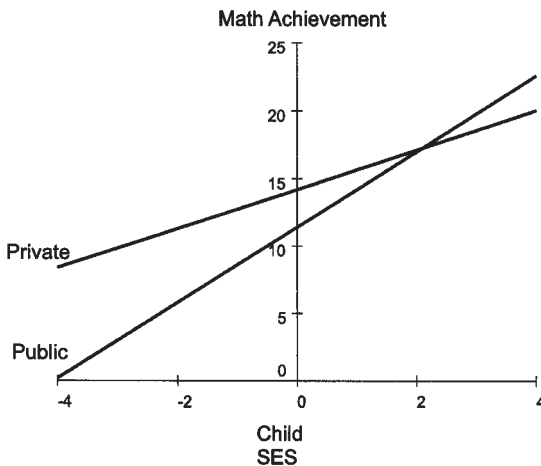


FIGURE 3 Plot of simple slopes between math achievement and child SES as a function of private versus public school. *Note.* The simple slope between math achievement and child SES significantly differs from zero within both private and public schools.

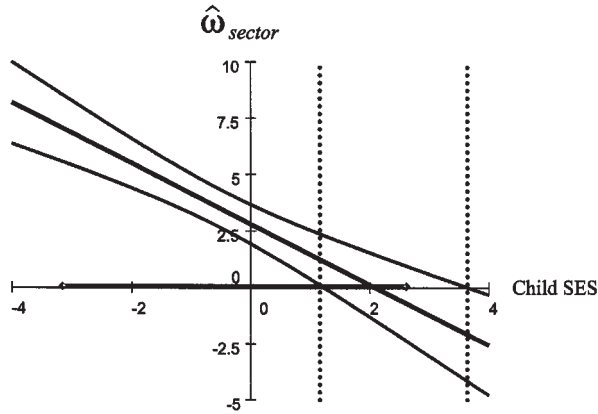


FIGURE 4 J-N regions of significance and confidence bands for the conditional relation between math achievement and child SES as a function of school sector. *Note.* Dashed vertical lines reflect regions of significance and dark horizontal line with diamonds indicates the actual range of child SES observed in sample data.

nificant at higher values of SES. Although the regions implied that the difference between private and public schools reversed direction at values of SES greater than 3.65, no cases were observed above this value in the HSB sample. The center of accuracy for estimating the achievement differences between students of private versus public schools was found at an SES level of $-.49$, with diminishing precision as SES increased or decreased from this value.

Child SES and school discipline. To demonstrate these procedures with a continuous by continuous interaction, we estimated a second multilevel model in which child SES remained the sole level-1 predictor, but the continuous measure of school discipline was considered as the sole level-2 predictor. Detailed results are presented in Table 3. The fixed effect for the cross-level interaction between child SES and school discipline was significant reflecting that the magnitude of the relation between child SES and math achievement varied across continuous levels of school discipline.

To probe the nature of this relation, we first used the pick-a-point approach in which we calculated the simple slopes between math achievement and child SES at high, medium, and low values of school discipline (defined as plus and minus one standard deviation around the mean of school discipline; see Figure 5). There was a significant and positive relation between child SES and math achievement across all three levels of school discipline, although the magnitude of this relation was larger at higher levels of school discipline (reflecting greater school discipline problems). Thus, child SES is a significantly stronger predictor of math achievement in schools with greater discipline problems.

TABLE 3
Results From HSB Multilevel Models With Continuous
by Continuous Interaction

<i>Model 2: Child SES and Disciplinary Climate</i>				
	<i>PE</i>	<i>SE</i>	<i>df</i>	<i>p</i>
Random effects				
Residual ($\hat{\sigma}^2$)	36.69	.63		< .0001
Intercept ($\hat{\tau}_{00}$)	6.64	.85		< .0001
Slope ($\hat{\tau}_{11}$)	.42	.25		.0459
Covariance ($\hat{\tau}_{01}$)	.84	.34		.0149
Fixed effects				
Intercept ($\hat{\gamma}_{00}$)	12.79	.22	158	< .0001
Child SES ($\hat{\gamma}_{10}$)	2.16	.12	7023	< .0001
Sector ($\hat{\gamma}_{01}$)	-1.49	.22	158	< .0001
Child SES by Sector ($\hat{\gamma}_{11}$)	.60	.13	7023	< .0001
-2LL			46651.6	
<i>Asymptotic Variance/Covariance Matrix of Fixed Effects</i>				
	$\hat{\gamma}_{00}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{01}$	$\hat{\gamma}_{11}$
$\hat{\gamma}_{00}$.048			
$\hat{\gamma}_{10}$.005	.015		
$\hat{\gamma}_{01}$	-.005	-.001	.050	
$\hat{\gamma}_{11}$	-.001	-.002	.006	.017

Note. All models were fit using SAS PROC MIXED using the High School and Beyond data described in the text. PE = parameter estimate; SE = standard error.

As before, the pick-a-point approach shows that there is a significant and positive relation between math achievement and child SES at high, medium, and low values of discipline and that the magnitude of this effect decreases with improved school discipline; however, we do not yet know *at what level* of school discipline the relationship between math achievement and SES becomes nonsignificant. Application of the J-N technique (see Figure 6), shows that the conditional effect of child SES on math achievement was significantly negative at discipline levels less than -6.38, nonsignificant at discipline levels between -6.38 and -2.47, and significantly positive at discipline levels above -2.47. Given that the range of observed values of discipline was between -2.28 and 2.89, this implies that there is a significant and positive relation between math achievement and child SES across all levels of school discipline observed within the HSB sample. The confidence bands show that the conditional effect of SES was estimated with most precision at a discipline level of .09, or roughly at the mean level of school discipline, with decreasing precision at higher or lower levels of school discipline. Again, the interested reader can reproduce these analyses, or conduct similar analyses of their own, through the freely accessible javascript programs described earlier.

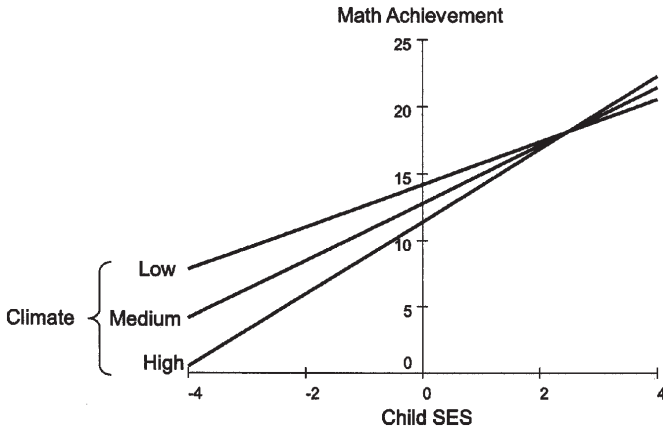


FIGURE 5 Plot of simple slopes of the relation between math achievement and child SES as a function of high, medium and low values of school disciplinary climate. *Note.* High, medium, and low values of school disciplinary climate are defined as plus and minus 1 *sd* about the mean.

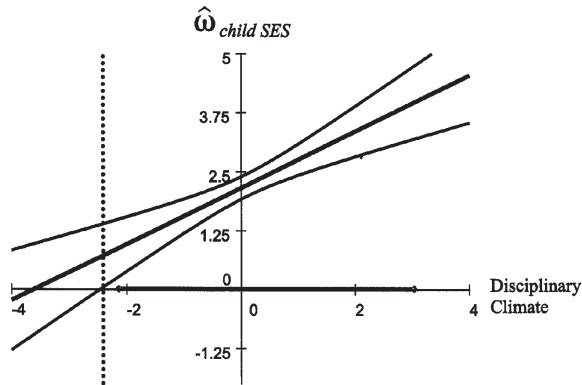


FIGURE 6 J-N regions of significance and confidence bands for the conditional relation between math achievement and child SES across all possible values of school disciplinary climate. *Note.* Only the upper boundary of the region is demarcated with the vertical dashed line given the scaling of the abscissa. Dark horizontal line with diamonds indicates the actual range of child SES observed in sample data.

LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

By highlighting the common analytical basis of the pick-a-point approach and J-N technique for evaluating conditional effects in models including interactions, we have shown that the J-N technique can be generalized to interactions involving any combination of categorical and continuous predictors, and that both the

pick-a-point and J-N techniques can be extended from fixed-effects regression models to multilevel models with random effects. We believe that these techniques provide critically important information needed to gain a full understanding about the complex conditional relations commonly encountered in both the fixed-effects and multilevel regression models. Although we believe that we have delineated the use of pick-a-point and J-N approaches in ways not previously considered, there are of course several limitations to our work.

One limitation of the present research is that we have not considered how the violation of specific model assumptions might impact on the use of either the pick-a-point approach or the J-N technique. We can, however, make several tentative statements on the basis of prior research. First, if the variance components of the model are not correctly specified (either in the fixed or multilevel cases, e.g., through the omission of a random effect, failure to model heteroscedasticity or autocorrelation, or in assumptions concerning the distributions of the errors) this is likely to adversely affect the estimation of the asymptotic covariance matrix of the fixed effects estimates, in turn compromising the inferential tests provided by pick-a-point or the J-N technique. A second assumption of both the fixed- and multilevel regression models is that the predictors in the model are nonstochastic, meaning that they are known and fixed values measured without error. Of course, this assumption will rarely hold in practice. Fortunately, research by Rogosa (1977, pp. 94–95) on the J-N technique in fixed-effects regression models shows that the results of the J-N technique are robust to the use of stochastic predictors. If the predictors are error-free, the Type I error rate is unaffected, although power diminishes. Measurement error predictably causes the region of significance to shrink (and confidence bands to broaden), yet Type I errors may also occur through a displacement of the region of significance (see Rogosa, 1977, p. 78, for further detail). Although clearly speculative, we would expect that similar results would also obtain for the J-N technique in multilevel models with stochastic predictors. Future research on the consequences of violating these and other assumptions of the fixed and multilevel regression models would be useful for ascertaining the robustness of the pick-a-point approach and J-N technique. Again, given that some assumptions will likely be violated in any given application, we believe it best to consider the results of the J-N technique as heuristic rather than exact.

A second limitation of the present research is that we discussed only two-way interactions between a single pair of predictors. The application of these techniques to more complex interaction patterns may at times be difficult. For instance, suppose that the focal predictor x_1 independently interacts with both with x_2 and with x_3 (i.e., there is a two-way interaction between x_1 and x_2 and a two-way interaction between x_1 and x_3). Then the conditional effect ω_1 is no longer a linear function of one moderator but is instead described by a *plane* over the dimensions x_2 and x_3 . The confidence bands then evolve into confidence sheets about the plane and the regions are correspondingly more complex to derive and interpret. In the ANCOVA context, Hunka (1995) and Hunka and Leighton (1997) discussed ways

to make the J-N technique tractable when there are multiple covariates that each interact with the grouping variable, but this approach has yet to be generalized to continuous focal predictors or models involving random effects. Similarly, the J-N technique may be difficult to apply in models involving three-way interactions, such as where x_1 interacts with the x_2x_3 product term. For situations such as these that may arise in the multilevel growth model, Curran et al. (2004) suggested it may be fruitful to blend the pick-a-point approach and J-N technique. Specifically, the conditional effect of x_1 would be plotted as a function of x_2 and examined through the J-N technique at various selected levels of x_3 . This approach has particular appeal if at least one of the moderators is nominal, providing natural levels at which to assess the conditional effects of the others. Further research may offer additional opportunities to explore higher-order interactions in fixed-effects and multilevel regression models.

REFERENCES

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Aitken, M. A. (1973). Fixed-width confidence intervals in linear regression with applications to the Johnson-Neyman technique. *British Journal of Mathematical and Statistical Psychology*, *26*, 261–269.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Bauer, D. J., Curran, P. J., & Bollen, K. A. (2001, June). *On the use of confidence bands in latent trajectory models*. Paper presented at the meeting of the Psychometric Society, King of Prussia, PA.
- Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158.
- Bryk, A. S., & Thum, Y. M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal*, *26*, 353–383.
- Cohen, J. (1978). Partialled products are interactions; Partialled powers are curve components. *Psychological Bulletin*, *85*, 858–866.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods*, *9*, 220–237.
- Curran, P. J., Bauer, D. J., & Willoughby, M. T. (in press). Testing and probing within-level and between-level interactions in hierarchical linear models. In C. S. Bergeman & S. M. Boker (Eds.), *The Notre Dame Series on quantitative methodology, Volume 1: Methodological issues in aging research*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Gafarian, A. V. (1964). Confidence bands in straight line regression. *Journal of the American Statistical Association*, *59*, 182–213.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Huitema, B. E. (1980). *The analysis of covariance and alternatives*. New York: Wiley.
- Hunka, S. (1995). Identifying regions of significance in ANCOVA problems having non-homogeneous regressions. *British Journal of Mathematical and Statistical Psychology*, *48*, 161–188.

- Hunka, S., & Leighton, J. (1997). Defining Johnson-Neyman regions of significance in the three-covariate ANCOVA using Mathematica. *Journal of Educational and Behavioral Statistics, 22*, 361–387.
- Hussong, A. M. (2003). Further refining the stress-coping model of alcohol involvement. *Addictive Behaviors, 28*, 1515–1522.
- Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.
- Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.
- Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika, 15*, 349–367.
- Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs, 1*, 57–93.
- Kacker, R. N., & Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association, 79*, 853–862.
- Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963–974.
- Miyazaki, Y. (2002, April). *Johnson-Neyman type technique in hierarchical linear model*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.
- Morrison, D. F. (1990). *Multivariate statistical methods*. New York: McGraw-Hill.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Boston: McGraw-Hill.
- Potthoff, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika, 29*, 241–256.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.
- Rogosa, D. (1977). Some results for the Johnson-Neyman technique. *Dissertation Abstracts International, 38* (09), 5366A. (UMI No. AAT 7802225).
- Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin, 88*, 307–321.
- Rogosa, D. (1981). On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within group regressions. *Educational and Psychological Measurement, 41*, 73–84.
- SAS Institute. (1999). *SAS documentation, Version 8*. Cary, NC: SAS Publications.
- Schaalje, G. B., McBride, J. B., & Fellingham, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models using SAS Proc MIXED. *Journal of Agricultural, Biological, and Environmental Statistics, 7*, 512–524.
- Serber, G. A. F. (1977). *Linear regression analysis*. New York: Wiley.
- Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24*, 323–355.
- Tate, R. (2004). Interpreting hierarchical linear and hierarchical generalized models with slopes as outcomes. *The Journal of Experimental Education, 73*, 71–95.
- Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.
- Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology, 10*, 395–426.
- Yip, T., & Fuligni, A. J. (2002). Daily variation in ethnic identity, ethnic behaviors, and psychological well-being among American adolescents of Chinese descent. *Child Development, 73*, 1557–1572.