Organizational Research Methods
Volume 9 Number 1
January 2006 5-28
© 2006 Sage Publications
10.1177/1094428105283190
http://orm.sagepub.com
hosted at
http://online.sagepub.com

The Adequacy of Repeated-Measures Regression for Multilevel Research

Comparisons With Repeated-Measures ANOVA, Multivariate Repeated-Measures ANOVA, and Multilevel Modeling Across Various Multilevel Research Designs

Vilmos F. Misangyi
University of Delaware
Jeffery A. LePine
James Algina
University of Florida
Francis Goeddeke Jr.
New Baltimore, Michigan

The authors assess the suitability of repeated-measures regression (RMR) to analyze multilevel data in four popular multilevel research designs by comparing results of RMR analyses to results of analyses using techniques known to produce correct results in these designs. The findings indicate that RMR may be suitable for only a small number of situations and that repeated-measures ANOVA, multivariate repeated-measures ANOVA, and multilevel modeling may be better suited to analyze multilevel data under most circumstances. The authors conclude by offering recommendations regarding the appropriateness of the different techniques given the different research designs.

Keywords: *repeated-measures regression; ANOVA; MANOVA; multilevel modeling; multilevel research*

Interest in research that acknowledges multiple levels of theory and analysis has been on the increase in organizational research (e.g., Chan, 1998b; Klein & Kozlowski, 2000; Klein, Tosi, & Cannella, 1999). This increase in multilevel research has given rise to analytical techniques specifically suited for multilevel modeling (Klein & Kozlowski, 2000), as well as greater use of repeated-measures analyses (Bergh, 1995). In fact, a review of three major journals in which organizational researchers publish (*Journal of Applied Psychology, Per-*

Authors' Note: Please address correspondence to Vilmos F. Misangyi, Department of Business Administration, Alfred Lerner College of Business & Economics, University of Delaware, Newark, DE 19716; e-mail: vilmos@udel.edu.

sonnel Psychology, and *Organizational Behavior and Human Decision Processes*) revealed that of the 677 articles published between 2000 and 2003, 13% involved some type of multilevel modeling and 76% of these studies used some form of repeated-measures analysis (repeated-measures ANOVA, multivariate repeated-measures ANOVA, or repeated-measures regression). In general, the rise in interest in multilevel modeling over the past decade has resulted not only in an increase in the number of studies using conventional repeated-measures ANOVA but also in the use of repeated-measures regression (RMR). RMR has been largely the method of choice for researchers performing multilevel studies investigating decision making in team contexts (e.g., Beersma et al., 2003; Hollenbeck, Ilgen, LePine, Colquitt, & Hedlund, 1998; Hollenbeck, Ilgen, & Sego, 1994; Hollenbeck et al., 1995; Marks, Zaccaro, & Mathieu, 2000; Phillips, 2001). RMR has also been used to address other substantive questions in multilevel research. For example, Ryan, McFarland, Baron, and Page (1999) used RMR to assess how much variance in staffing practices was attributable to each of national and cultural differences.

Despite this use of RMR, there are few lucid examinations of the technique's efficacy for multilevel designs (Gully, 1997; Lorch & Myers, 1990), especially as compared to that of repeated-measures ANOVA (for instance, see Keselman, Algina, & Kowalchuk, 2001) or multilevel modeling (MLM) techniques such as hierarchical linear modeling (HLM; for instance, see Hofmann, Griffin, & Gavin, 2000). The rise of RMR in multilevel research is enough in itself to warrant a further investigation of its utility for such purposes. But such an examination is especially pertinent given the technique's apparent simplicity. RMR is relatively simple to learn: It is an extension of familiar repeated-measures ANOVA and builds on standard regression analyses (by manually adjusting results to reflect the relevant levels of analysis). Yet at the same time, RMR appears to be similar to much more sophisticated MLM techniques (e.g., HLM). RMR explicitly partitions and models within and between variance and also appears to be appropriate for continuous as well as categorical data (Cohen & Cohen, 1983; Gully, 1994, 1997; Hollenbeck et al., 1994; Pedhazur, 1982). Thus, it appears that RMR is the best of both worlds: Similar to the more sophisticated techniques, it explicitly takes into account nesting of the data at multiple levels and apparently can handle continuous variables. At the same time, it also maintains the simplicity of the more familiar approaches. If this latter point proves to be the case, then although newer, more complex statistical techniques may be tailored more specifically to analyze multilevel data, there is the question of whether such techniques outperform this simpler approach that is easier to learn and use.

To accomplish our goal of assessing the suitability of RMR to analyze multilevel data, we consider a subset of straightforward multilevel designs that highlight issues appropriate to the comparison of RMR relative to other techniques that are known to be correct given the relevant design. In making these assessments, we give priority to the adequacy of parameter estimates and statistical tests. When there are no differences across the techniques, we consider simplicity and clarity of presentation in our assessment. Thus, the practical contribution of this research is that we are able to offer guidance concerning when RMR may be most and least appropriate when conducting multilevel research. Of course, this approach also allows us to compare results of the different analytic techniques in each of the multilevel research designs, and as a consequence, we can offer recommendations regarding which technique is best suited for each design. Therefore, there is a broader practical contribution of this research beyond just an assessment of RMR, and we expect that our findings will complement other treatments of the issue (e.g., Klein et al., 2000).

To facilitate our assessment of RMR, we intentionally examine four organizationally relevant scenarios in their simplest form: (a) a repeated-measures design in which there is one within-subjects factor, (b) a between-groups design in which the only independent variable is group membership, (c) a between-groups design with one continuous group-level independent variable, and (d) a design in which there is a single continuous individual-level variable. The examination of each research design at a basic level allows for the focus on the problems/ capabilities of each analytical technique examined for the particular design. RMR is compared to repeated-measures ANOVA and multivariate repeated-measures ANOVA in the first design, as the properties of these techniques are well known for such a design. We also examine what, if any, benefits MLM analyses bring to such a simple design, thereby providing a more specific complement to the more general design provided by Hox's (2002) comparison of multivariate ANOVA and MLM for a design with within-subjects and between-subjects factors. RMR is compared to repeated-measures ANOVA and MLM in the second design, again given the known properties of the latter techniques. We then assess RMR's capabilities for the final two designs as compared to the known capabilities of MLM for these designs. Recommendations are provided at the end of each section. Finally, we conclude with an overview of the findings and a discussion of their implications for the application of RMR to more complex designs involving more than one continuous variable.

Four Cases of Multilevel Designs

Multilevel Designs With One Within-Group Factor

We first consider a straightforward repeated-measures design in which there is one within-subjects factor. In an organizational context, such a design might be used to assess the effects of individual job experience on individual job performance. We examine several alternative analytic techniques that can be used for this design: repeated-measures ANOVA, multivariate repeated-measures ANOVA, RMR, and MLM. Considering this design is important, not only because of its familiarity and usefulness to organizational scholars but also because it allows us to illustrate the effect, even in a very simple design, of a violation of the sphericity assumption.

The sphericity assumption is an assumption made in several of the analytic techniques under consideration and is relevant to the pattern of the variances and covariances for the scores obtained under the various treatments. The assumption implies that the covariance for each pair of treatments is a function of three parameters: the variance for the first repeated measure in the pair, the variance for the second repeated measure in the pair, and a parameter that is common to all covariances:

$$\sigma_{tt'} = \left(\sigma_t^2 + \sigma_{t'}^2\right) / 2 - \lambda. \tag{1}$$

In Equation 1, the subscripts t and t' denote any two treatments. When the sphericity assumption is not satisfied, inferences using procedures that make the sphericity assumption will be incorrect. This is especially pertinent when examining questions such as the effect of job experience on job performance. For example, Chan (1998a) highlights how the spheric-

Table 1
ANOVA Results for Repeated-Measures Design With One Within-Subjects Factor

Source	*df*	SS	MS	*F*	*p*
Treatment	6	0.07504791	0.01250799	3.23	.0072
Participants	12	0.57720376	0.04810031		
Error	72	0.27857009	0.00386903		

Note: *df* = degrees of freedom; SS = sum of squares; MS = mean squares.

ity assumption is most likely violated when considering intraindividual change over time. Let $C_{t+1,t} = X_{t+1} - X_t$ be the change score for time point t and the next time point, that is, time point $t + 1$. Although the sphericity assumption implies that the variance of $C_{t+1,t}$ is the same for all such pairs of time points, individuals may systematically differ in their individual growth parameters. In other words, it is reasonable to expect that the rate of change in performance over time will differ across individuals, and thus the variance of the change scores will also increase over time. When using ANOVA, the use of an F statistic that does not adjust the degrees of freedom for violation of this assumption results in a Type I error rate that is larger than the α used as a comparison for the p value. Consequently, the null hypothesis will be rejected more frequently than it should be (Box, 1954).

ANOVA. Table 1 contains the results of an ANOVA of data collected in a design with one within-subjects factor with $T = 7$ repeated measurements of $n = 13$ participants. For instance, if examining the effect of individual job experience on job performance, this data structure is similar to collecting a measure of the job performance for 13 different individuals over 7 relevant intervals of time (i.e., month or year). The p value is calculated as the area beyond $F = 3.23$ in an F distribution with $T - 1 = 6$ and $(n - 1)(T - 1)$ degrees of freedom, apparently indicating a significant effect of treatment ($p < .01$). However, as just discussed, these degrees of freedom are correct only if the data meet the sphericity assumption. When the sphericity assumption is not correct, an adjusted F test, such as the Huynh and Feldt (HF; 1976) ε-approximate F test, must be used. This adjustment corrects the degrees of freedom, using Box's epsilon, ε, such that the degrees of freedom are $\varepsilon(T - 1)$ and $\varepsilon(n - 1)(T - 1)$. In general, ε must be between $1/(T - 1)$ and 1.0, with values near 1.0 indicating less severe violations of sphericity.

For the current data, the HF estimate of ε is $\tilde{\varepsilon} = .2361$ (which is near the lower limit of 1/6), suggesting a violation of the sphericity assumption. (See Huynh & Feldt, 1976, for the formula for $\tilde{\varepsilon}$.) The estimated HF approximate degrees of freedom are $.2361 \times 6 = 1.42$ and $.2361 \times 72 = 17.00$. The corrected estimate of the p value is the area beyond 3.23 in the F distribution with 1.42 and 17.00 degrees of freedom. This is $p = .0719$, and the treatment effect is not significant. It should be noted that ANOVA and general linear modeling (GLM) procedures in programs such as SAS and SPSS routinely calculate the corrected p value. Thus, ANOVA is a viable method for this simple design, as long as the proper adjustments are made when the sphericity assumption is violated.

It is worth noting that it is well known that under the assumption of multivariate normality, the sphericity assumption can be tested by using the likelihood ratio test (see, e.g., Hox, 2002). The likelihood ratio is the basis for the Mauchly test produced in SAS and SPSS. The availability of this test suggests the following strategy for using the test of the sphericity assumption: If the assumption is not rejected, use repeated-measures ANOVA without the adjusted degrees of freedom; if it is rejected, use repeated-measures ANOVA with the adjusted degrees of freedom (alternatively, one could use multivariate repeated-measures ANOVA, or MLM; see below). Keselman, Rogan, Mendoza, and Breen (1980) compared this strategy to always using repeated-measures ANOVA with the adjusted degrees of freedom and found that the two procedures have similar Type I error rates and power, so testing for sphericity seems to be unnecessary.

Multivariate repeated-measures ANOVA. An alternative to the just described univariate ANOVA approach is a multivariate repeated-measures ANOVA. For the current data, this approach yields $F(6, 72) = 58.58$, $p < .05$. This analysis is based on the assumption that the repeated measurements are drawn from a multivariate normal distribution and that scores for different participants are independent of one another. These assumptions are also made by all the other methods discussed in this section, including the ANOVA technique just discussed. The multivariate approach is not based on the sphericity assumption, however. Instead, this approach makes no assumption about the variances and covariances for the scores obtained under the various treatments. The variances and covariances are said to be unstructured. The multivariate approach applied to the data indicates a significant treatment effect, whereas the adjusted univariate test did not. These results are consistent with those of Algina and Keselman (1997), who provide evidence that the multivariate approach will have greater power than that for the univariate approach when ε is small. Recall in the analysis above that $\tilde{\varepsilon} = .2361$ is relatively small, as it is near the lower limit of 1/6, suggesting a severe violation of sphericity. In sum, the validity of the F test in the multivariate analysis of repeated measures is not affected by whether the data meet the sphericity assumption and will often be more powerful than the ANOVA approach.

Repeated-measures regression. The RMR approach uses the following test statistic

$$F = \frac{(n-1)(T-1)}{T-1} \frac{R^2_{TMT}}{(1 - R^2_{BET} - R^2_{TMT})}, \tag{2}$$

under the assumption that the F statistic has degrees of freedom of $T - 1 = 6$ and $(n - 1)(T - 1) = 72$. In the test statistic, R^2_{TMT} is the squared multiple correlation for a model predicting scores on the dependent variable from a model with six dummy-coded variables indicating the treatment level in which the score was obtained and R^2_{BET} is the squared multiple correlation for a model predicting scores on the dependent variable from a model with 12 dummy variables indicating the person who contributed the score. The calculated value of the test statistic is

$$F = \frac{(13-1)(7-1)}{7-1} \frac{.0806}{(1-.6201-.0806)} = 3.23,$$

which is the same as obtained in the ANOVA. The R^2 figures used in the RMR can easily be obtained by using a regression analysis program; however, the dummy coding is more tedious and error prone than is using the ANOVA and GLM programs in commercially available packages. More important, regression analysis programs would not produce estimated ε or corrected p values, and therefore the RMR approach should be used only if one is confident

Table 2
Pairwise Comparisons of Treatment Effects With Assumptions Consistent to the Repeated-Measures Regression Approach

Pair	Mean Difference	Standard Error	t	p
2 vs. 1	0.07462	0.03959	1.88	.0629
3 vs. 1	0.03946	0.03959	1.00	.3217
4 vs. 1	0.07469	0.03959	1.89	.0627
5 vs. 1	0.09362	0.03959	2.36	.0204
6 vs. 1	0.05154	0.03959	1.30	.1965
7 vs. 1	0.03915	0.03959	0.99	.3255

Note: Results from normal ordinary least squares regression in which the standard errors are computed under the assumption that the scores contributed by a particular participant are independent.

Table 3
Pairwise Comparisons of Treatment Effects With Assumptions Consistent to the Multivariate Repeated-Measures ANOVA Approach

Pair	Mean Difference	Standard Error	t	p
2 vs. 1	0.0746154	0.0051348	14.53	<.0001
3 vs. 1	0.0394615	0.0138890	2.84	.0149
4 vs. 1	0.0746923	0.0274484	2.72	.0186
5 vs. 1	0.0936154	0.0375109	2.50	.0281
6 vs. 1	0.0515385	0.0378818	1.36	.1987
7 vs. 1	0.0391538	0.0325117	1.20	.2517

Note: The dependent sample t test only assumes that the repeated measurements are drawn from a multivariate normal distribution and that scores for different participants are independent of one another. Variances and covariances for the scores obtained under the various treatments are presumed to be unstructured.

that the sphericity assumption is met; otherwise, there is an increased probability of Type I error.

To further illustrate the implications of the assumptions in RMR, Tables 2 and 3 present the results (mean difference, standard error of the mean difference, t statistic, and p value) for a subset (i.e., as compared to Time Period 1) of the pairwise comparisons of the seven treatments tested using assumptions consistent with the RMR approach (Table 2) and with the multivariate repeated-measures ANOVA approach (Table 3). Table 2 shows the statistics from the model used to calculate R^2_{TMT} in the RMR approach, which are from a standard ordinary least squares regression and based on its pursuant assumptions (i.e., independence). For the multivariate repeated-measures ANOVA approach reported in Table 3, each t statistic is computed using the formula for a dependent sample t statistic,

$$t = \frac{\bar{X}_t - \bar{X}_{t'}}{\sqrt{\dfrac{S_t^2 + S_{t'}^2 - 2\,S_{t'}\,S_{t'}\,r_{tt'}}{n}}}, \tag{3}$$

Table 4
Pairwise Comparisons of Treatment Effects Consistent With the Repeated-Measures Regression Approach With the Cohen and Cohen (1983) Adjustment

Pair	Mean Difference	Standard Error	t	p
2 vs. 1	0.07461538	0.02439744	3.06	.0031
3 vs. 1	0.03946154	0.02439744	1.62	.1102
4 vs. 1	0.07469231	0.02439744	3.06	.0031
5 vs. 1	0.09361538	0.02439744	3.84	.0003
6 vs. 1	0.05153846	0.02439744	2.11	.0381
7 vs. 1	0.03915385	0.02439744	1.60	.1129

Note: In contrast to the dependent samples t test, the Cohen and Cohen (1983) correction is made under the sphericity assumption.

and is distributed on $n - 1$ degrees of freedom. In Equation 3, $\overline{X}_t$ and $\overline{X}_{t'}$ are means for job performance at times t and t', S_t^2 and $S_{t'}^2$ are variances for job performance at times t and t', and $r_{tt'}$ is the correlation between job performance at time t and t'. Because Equation 3 is the dependent samples t statistic, it can easily be computed in SAS or SPSS. These t statistics are based only on the assumption that the repeated measurements are drawn from a multivariate normal distribution and that scores for different participants are independent of one another. Similar to the assumptions of the multivariate repeated-measures ANOVA, the t statistics are not based on the sphericity assumption. Instead, the variances and covariances for the scores obtained under the various treatments are presumed to be unstructured.

A comparison of Tables 2 and 3 shows that the mean differences are the same for the two approaches, but the standard errors, t statistics, and p values are not. The standard errors are incorrect in Table 2 (consistent with the RMR approach) because they are computed under the assumption that the seven scores contributed by a particular participant are independent. As a consequence, the t statistics and p values are wrong.

Cohen and Cohen (1983, Equation 11.2.8) provided a correction for the t statistic,

$$\sqrt{\frac{T-1}{T} \times \frac{1-R_{TMT}^2}{1-R_{BET}^2-R_{TMT}^2}}\, t, \tag{4}$$

as well as a correction for the standard errors when using RMR. Corrected t statistics and standard errors are shown in Table 4. Comparing the results in Table 2 and Table 4 shows that the standard errors for a particular comparison are not equal, and as a result, the corresponding t statistics in the two tables are not equal, nor are the p values. This is because the results in Table 4, consistent with the RMR approach, are calculated under the sphericity assumption. Boik (1981) has shown that when the sphericity assumption is violated, these t statistics can result in very liberal or very conservative tests of the pairwise comparisons and should be avoided unless it is known that the sphericity assumption is not violated. Thus, these results further show that RMR should not be used in analyzing multilevel data unless one is confident that the sphericity assumption is satisfied in the data.

MLM. Finally, these repeated-measures data can also be analyzed using MLM. MLM is a very flexible approach to analyzing data and can be implemented in a number of ways. We first analyzed the data using an approach presented in Gulley (1997) in which the mean scores over time for individuals are assumed to vary across individuals but the time effects are assumed to be fixed:

$$Y_{it} = \gamma_{0i} + \gamma_1 z_2 + \gamma_2 z_3 + \gamma_3 z_4 + \gamma_4 z_5 + \gamma_5 z_6 + \gamma_6 z_7 + \varepsilon_{it}. \quad (5)$$

This model is called a random-intercepts model. The variables z_2 to z_7 are dummy coded variables with $z_j = 1$ for scores obtained under the treatment indicated by the subscript and zero otherwise. The parameter γ_{0i} is the intercept and is assumed to vary randomly over participants. The parameters γ_1 to γ_6 are assumed not to vary over participants. The variable ε_{it} is a residual assumed to vary randomly over participants and time and is not assumed to have the same variance at all time points. These assumptions about γ_{0i} and ε_{it} are equivalent to the sphericity assumption made in the ANOVA and RMR approach. Therefore, the F statistic for comparing the treatments is the same for this MLM model as it is for the RMR and ANOVA approaches. Moreover, this MLM approach does not incorporate the p value adjustment required for violation of sphericity, and the pairwise comparison t statistics associated with this approach are equivalent to the statistics obtained by Cohen and Cohen's (1983) correction.

An alternative MLM takes the form of

$$Y_{it} = \gamma_{0i} + \gamma_{1i} z_2 + \gamma_{2i} z_3 + \gamma_{3i} z_4 + \gamma_{4i} z_5 + \gamma_{5i} z_6 + \gamma_{6i} z_7 + \varepsilon_{it}. \quad (6)$$

Here, the i subscripts on the coefficients of the dummy variables indicate the assumption that the time effects, as well as the intercepts, vary across participants (random effects modeling). The model is called a random-coefficients regression model. If we assume that all of the γ coefficients are correlated, the variances and covariances for the scores obtained under the various treatments are presumed to be unstructured. Then, using the Hotelling-Lawley-McKeon F statistic, available in PROC MIXED in SAS, this MLM analysis can be used to obtain an F statistic that is equal to the statistic obtained by a multivariate repeated-measures ANOVA of the data. Moreover, the t statistics we obtain for pairwise comparisons are equal to those in Table 3 and are based only on the assumption that the repeated measurements are drawn from a multivariate normal distribution and that scores for different participants are independent of one another. According to Wolfinger and Chang (1995), the Hotelling-Lawley-McKeon F statistic performs better than the default F statistic that is available in SAS and is used in most MLM programs.

Although it appears that MLM has little utility for a design with one repeated-measures factor, because it produces results identical to the multivariate repeated-measures ANOVA approach, the flexibility of the MLM approach does offer two advantages over the other methods in certain situations. First, users of MLM can make other assumptions besides the sphericity and unstructured assumption. If a correct or nearly correct assumption is made when appropriate, the analysis will tend to have more power than if it were based on either the sphericity assumption or the assumption that the variances and covariances are unstructured. Second, with missing data, MLM will use all of the available data and, with some types of missing data, will provide valid tests of the hypotheses whereas the use of ANOVA, multivariate repeated-measures ANOVA, and RMR become problematic when there is missing data (see Keselman, Algina, Kowalachuk, & Wolfinger, 1999) for a recent discussion of both of these benefits of using MLM in the context of a repeated measures design).

Summary. We summarize this section regarding simple repeated-measure designs with one within-subjects factor with the following observations:

1. The use of ANOVA to analyze the data will produce an incorrect F test if the data do not meet the sphericity assumption. The analysis can be easily corrected, however, using an estimate of Box's ε, which is routinely produced by commercially available statistical software.
2. The validity of the F test in the multivariate repeated-measures ANOVA is not affected by whether the data meet the sphericity assumption and will often be more powerful than the univariate ANOVA approach.
3. The RMR analysis of the data produces the same F statistic as produced by the ANOVA, as it is also subject to the sphericity assumption. The estimate of Box's ε is not available in commercial statistical packages, however. Furthermore, pairwise comparison t tests produced using Cohen's correction will be incorrect if the sphericity assumption is violated.
4. A random-intercept MLM produces results that are the same as those produced by RMR. A random-coefficients MLM approach, however, can produce an F statistic that is equal to the F statistic for the multivariate repeated-measures analysis. This latter MLM also offers researchers the ability to make alternative assumptions regarding the variance-covariance matrix beyond the limited assumptions of the other methodologies as well as can handle the presence of missing data.
5. In sum, given that the assumption of sphericity is often likely to be violated in organizational research (Chan, 2002), it appears that multivariate repeated-measures ANOVA or MLM procedures offer the best alternatives for this simple repeated-measures design. Although there does not seem to be any reason to use the RMR approach for this research design, the use of ANOVA can be an effective analytic technique when the user is confident that the sphericity assumption is satisfied or, if this assumption is not met, the appropriate adjustments are made. The multivariate repeated-measures ANOVA, however, is an alternative that can be more powerful than ANOVA when the sphericity assumption is severely violated by the data. MLM is an equivalent alternative to the multivariate repeated-measures ANOVA when the Hoteling-McKeon-Lawley F statistic is used, unless one is confident that an alternative assumption about the variances and covariances should be made or if there are missing data. If either or both of these situations are presented by the data, then MLM is a better approach than is the multivariate repeated-measures ANOVA. MLM is particularly advantageous when there are missing data, and in this situation, improved performance of the F test of equality of the means on the repeated measures is obtained by using the Kenward-Roger test available in SAS (see Padilla & Algina, 2004).

Multilevel Designs With Group Membership as the Only Between-Groups Factor

We next examine a design that is used to assess the extent to which the group mean varies across groups. Such a design is highly relevant in the context of research focused on establishing the construct validity of a group-level variable. For the examination of the utility provided by the various analytic techniques for this and the subsequent research design, we use a subset of the simulated data provided by Klein and her colleagues (2000). The analysis presented in Klein et al. (2000) of these data includes hypotheses regarding the relationships that several individual- and group-level variables have with job satisfaction and thus is well suited for the current purposes. See Table 5 for information on the subset of the simulated data.

Table 5
Descriptive and Summary Statistics and Correlations for Selected Variables From the Klein et al. (2000) Simulated Data Set

Variable	Descriptive and Summary Statistics[a] *M*	*SD*	ICC (1)	ICC (2)
Job satisfaction	0.08	2.48	.11	.65
Pay	0.04	1.01	–.01	–.27
Negative leadership	0.00	1.09	.04	.38

Variable	Correlations Between Independent Variables and Job Satisfaction[b] Raw (n = 750)	Between-Group (n = 50)	Within-Group (n = 700)
Pay	0.37*	–0.08	0.42*
Negative leadership	–0.29*	–0.61*	–0.25*

Note: ICC = intraclass correlation coefficient.
a. Excerpted from Table 12.3 of Klein et al. (2000).
b. Excerpted from Table 12.4 of Klein et al. (2000).
*p < .01, two-tailed.

Table 6
ANOVA Results of Klein et al. (2000) Data: Group Membership as the Only Between-Groups Factor

Source	*df*	SS	MS	EMS	*F*	*p*
Groups	49	770.622067	15.726981	$\sigma_W^2 + n\sigma_B^2$	2.87	<.0001
Error	700	3831.886397	5.474123	σ_W^2		
Total	749	4602.508464				

Note: *df* = degrees of freedom; SS = sum of squares; MS = mean squares; EMS = expected mean squares.

With this nested design, the total variance in job satisfaction is conceptualized as comprising two components: between-group variance (σ_B^2) and within-group variance (σ_W^2). This partitioning of the variance enables two useful analyses. First, it allows us to estimate the intraclass correlation coefficient (ICC),

$$\mathrm{ICC} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}. \tag{7}$$

The ICC is a measure of the proportion of total variance in job satisfaction that is between groups; it is the amount of total variance attributable to group membership. A second purpose is to test the null hypothesis that $\sigma_B^2 = 0$. Rejection of the null hypothesis implies that group-mean job satisfaction varies across groups. This test has been used as a hurdle to support the construct validity of group-level variables (Chan, 1998b). For the current research design, we examine several analytic techniques: ANOVA, RMR, and MLM.

ANOVA. A summary of an ANOVA of the Klein et al. (2000) data is presented in Table 6 along with the expected mean squares (EMS) for the groups and error sources of variance.

The EMS values are correct for a balanced design, that is, a design in which the number of data points within each group is constant across groups. Based on the EMS, we can estimate the within-groups variance as 5.474 and the between-groups variance as

$$\hat{\sigma}_B^2 = \frac{MS_{Groups} - MS_{Error}}{n} = \frac{15.726 - 5.474}{15} = .683. \tag{8}$$

The estimated ICC is

$$\hat{ICC} = \frac{.683}{.683 + 5.474} = .112,$$

indicating that about 11% of the total variance is between groups. The statistic $F(49, 700) = 2.87$, $p < .0001$, indicates that the null $\sigma_B^2 = 0$ can be rejected.

It might be noted that if the data are not balanced, then

$$\hat{\sigma}_B^2 = \frac{MS_{Groups} - MS_{Error}}{n_0}, \tag{9}$$

where

$$n_0 = \frac{1}{G-1}\left(N - \frac{\sum_{g=1}^{G} n_g^2}{N} \right) \tag{10}$$

and n_g is the frequency in the gth group.

RMR. The Klein et al. (2000) data can also be analyzed by using a form of RMR. The regression analysis can be conducted by using the model

$$Y_{ig} = \alpha + \beta_1 z_1 + \ldots + \beta_{49} z_{49} + \varepsilon_{ig}, \tag{11}$$

where the z variables are dummy-coded variables, with $z_j = 1$ for scores by the participant indicated by the subscript and zero otherwise. Instead of an ICC, this analysis estimates the squared multiple correlation coefficient, which in this case is .167. This statistic can be calculated by using

$$R^2 = \frac{SS_{Group}}{SS_{Error} + SS_{Group}}, \tag{12}$$

where the sums of square are obtained from the ANOVA of the same data. In other words, RMR estimates that 16.7% of the total variance in job satisfaction is between groups compared to the 11.2% ICC estimate in ANOVA.

Using the EMS expressions from the ANOVA, it can be shown that the ratio of the quantity estimated by the numerator to the quantity estimated by the denominator in the R^2 estimate of RMR is

$$R^2 = \frac{\frac{\sigma_W^2}{n} + \sigma_B^2}{\frac{(N-1)}{N-n}\sigma_W^2 + \sigma_B^2}. \tag{13}$$

This expression suggests that R^2 using the RMR approach is a deficient estimator of the ICC. Because of the term σ_W^2/n in the numerator, R^2 will tend to be too large when the within-group variance is large and the number of participants in a group is small. An alternative to R^2 is the adjusted R^2,

$$R_a^2 = \frac{N-1}{N-G}R^2 - \frac{G-1}{N-G} = \frac{750-1}{750-50}.167 - \frac{50-1}{750-50} = .109, \tag{14}$$

which for the Klein et al. (2000) data is closer to $\hat{\text{ICC}} = .112$, but still not correct. The estimators $\hat{\text{ICC}}$ and R_a^2 can be substantially different in small samples, and thus $\hat{\text{ICC}}$ is regarded as the better estimator of the *ICC*. The *F* statistic based on the regression analysis, however, provides exactly the same result as the ANOVA:

$$F = \frac{N-G}{G-1}\frac{R^2}{1-R^2} = \frac{750-50}{50-1}\frac{.163}{1-.163} = 2.87. \tag{15}$$

The implication, then, is that although RMR may correctly assess whether there is significant between-group variance, it produces inferior estimates of the proportion of total variance that is due to the between-group variance.

MLM. Using the MLM approach, the variances σ_B^2 and σ_W^2 are typically estimated by using either maximum likelihood estimation or restricted maximum likelihood estimation (RMLE); RMLE seems to be preferred for this design (see McCulloch & Searle, 2001). When the groups are equal in size, as in the current example, the RMLE estimate of σ_W^2 is equal to the ANOVA estimate. The RMLE estimate of σ_B^2 is equal to the ANOVA estimate, provided that this estimate is not negative. If the latter estimate is negative, the RMLE estimate of σ_B^2 is zero. Of course, if the ANOVA estimate of σ_B^2 is negative, it is common to report that the between-groups variance is zero, so that in practice, when the groups are equal in size, the ANOVA and RMLE estimates are equal. If the data are not balanced (i.e., the groups are not equal in size), the RMLE and ANOVA estimates are not equal. The RMLE estimates are preferable (McCulloch & Searle, 2001).

When MLM is used, there are several methods for testing the null hypothesis that $\sigma_B^2 = 0$ (see Berkhof & Snidgers, 2001, for a description of the methods), and it is possible that the decisions about the null hypothesis will vary across the available MLM-based procedures for testing the hypothesis. Furthermore, because MLM procedures apply asymptotic tests, they may not be appropriate for small sample sizes.

Summary. In summary, for designs in which the only between-groups factor is information on group membership, we found the following:

1. With regard to the estimation of σ^2_B and σ^2_W, for balanced designs, the ANOVA estimates can be used. There is no advantage to MLM analyses using RMLE estimates over the ANOVA estimates. For unbalanced designs, however, the MLM approach using the RMLE estimates should be used.
2. The test of the hypothesis $\sigma^2_B = 0$ is best tested by the ANOVA F test. For both balanced and unbalanced designs, the ANOVA test is preferable to the tests associated with MLM because the former test is correct regardless of the sample size. The MLM tests are asymptotic tests and can provide incorrect results for small sample sizes.
3. Overall, there is no reason to use RMR to analyze the data given this type of design. RMR produces biased estimates of the amount of variance attributed to group membership. The F test is the same as the ANOVA F test and thus offers no advantages.

Multilevel Designs With One Continuous Group-Level Independent Variable

The third design we consider is one in which a single continuous group-level variable is related to an individual-level dependent variable. Such a design is useful, for example, to answer the question of how some group characteristic (e.g., cohesion) influences individual affect (e.g., satisfaction) or behavior (performance). Often, the group-level variable is the mean of group members' ratings on the relevant group characteristic, and support for using the mean to represent the group is normally a statistically significant ICC, as discussed earlier. The simulated data used by Klein et al. (2000) included data on employees' assessments of negative leader behaviors, and thus, one research question might be whether the group mean assessment of negative leader behaviors is related to individual job satisfaction. We use these data to compare RMR with MLM, given the known properties of the latter technique for such a design.

MLM. In the MLM literature, a means as outcomes model is used to test for this type of relationship. The Level 1 model is

$$Y_{ig} = \beta_{0g} + r_{ig} \tag{16}$$

and states that the job satisfaction score for employee i in group g is equal to the expected value of job satisfaction scores for employees in that group. The simultaneous Level 2 model is

$$\beta_{0g} = \gamma_0 + \gamma_1 \bar{X}_g + u_{0g}, \tag{17}$$

where $\bar{X}_g$ is the mean negative leader behaviors score for employees in group g. Equation 17 is a model for the relationship between mean job satisfaction in group g and mean negative leader behaviors score in group g. The parameter γ_1 is a regression coefficient relating the means on the two variables. Substituting Equation 17 into Equation 16, we obtain the combined model

$$Y_{ig} = \gamma_0 + \gamma_1 \overline{X}_g + u_{0g} + r_{ig}, \tag{18}$$

and we see that γ_1 is also a regression coefficient relating individual job satisfaction to the group mean on negative leader behaviors. The MLM estimate of γ_1 is –1.78, and the F statistic for testing the null hypothesis $\gamma_1 = 0$ is $F(1, 48) = 27.80, p < .0001$, supporting a relationship between job satisfaction and group mean negative leader behaviors.

RMR. An RMR analysis relies on the results from two separate regressions. The F statistic is

$$F = (G-2)\frac{R^2_{NLB}}{R^2_{Bet} - R^2_{NLB}}, \tag{19}$$

where R^2_{NLB} is the squared multiple correlation coefficient for the regression model

$$Y_{ig} = \alpha + \beta_1 \overline{X}_g + \varepsilon_{ig} \tag{20}$$

and R^2_{Bet} is the squared multiple correlation coefficient calculated in the previous section using Equations 11 and 12. Substitution in the formula for F yields

$$F = (50-2)\frac{.0614}{.1674 - .0614} = 27.80$$

and is equal to the F from the MLM. Like the MLM F, the RMR F has 1 and 48 degrees of freedom. Thus, for balanced designs, the MLM and RMR approach yield the same F test. It might be noted that the estimate of β_1 in Equation 20 is –1.78 and is exactly equal to the MLM estimate.

Unbalanced designs. If the research design is not balanced, the MLM and RMR approaches do not yield the same F statistic, although in our experiences, both F statistics will be quite similar. In addition, $\hat{\gamma}_1$ and $\hat{\beta}_1$ will tend to be similar. For example, we randomly deleted cases from the Klein et al. (2000) data with a .5 probability of deletion. In the resulting sample, group sizes ranged from 5 to 10. For MLM, $\hat{\gamma}_1 = -1.28$ and $F(1, 42.5) = 17.52, p < .0001$; for RMR, $\hat{\beta}_1 = -1.30$ and $F(1, 48) = 18.65, p < .0001$. However, the MLM analysis is the theoretically correct analysis, and the operating characteristics (Type I and Type II error rates) of the RMR analysis are unknown. Therefore, although both RMR and MLM can be used for balanced designs, MLM is preferable for unbalanced designs.

Summary. In summary, we observed the following for designs in which a continuous group-level variable was related to an individual-level dependent variable.

1. For balanced designs, MLM and RMR produce the same F value and can be used interchangeably.
2. For unbalanced designs, the F test for MLM and RMR will likely be similar; however, the MLM results are supported by statistical theory, and the operating characteristics of the RMR technique are unknown. Thus, the MLM approach is preferable for large samples. It is unclear

which approach is better for unbalanced designs using small samples: Although the asymptotic tests of MLM may produce incorrect results, the operating characteristics of the RMR technique are unknown.

Multilevel Designs With One Continuous Individual-Level Independent Variable

The final design we assess is one in which the focus is on the effect of individuals' standings on some continuous variable. For example, Klein et al. (2000) included data on employees' pay and examined whether it is related to employees' job satisfaction. We use these data to examine the effectiveness of the MLM and RMR approaches in analyzing this research design under two different assumption scenarios. In the first scenario, the relationship of job satisfaction to pay is the same for all groups. In the second, the relationship of job satisfaction to pay varies across groups.

The relationship of job satisfaction to pay is the same for all groups. Using MLM, the model for investigating this scenario is the random intercepts model and follows one of three alternative approaches. In the first "raw form" approach, the Level 1 model is

$$Y_{ig} = \beta_{0g} + \gamma_1 X_{ig} + r_{ig}, \tag{21}$$

and the Level 2 model is

$$\beta_{0g} = \gamma_0 + \gamma_2 \overline{X}_g + u_{0g}. \tag{22}$$

The coefficient γ_1 is the within-groups regression coefficient of the relationship of employees' pay to employees' job satisfaction, which is assumed to be the same for all groups. The coefficient β_{0g} is an intercept specific to group g and is assumed to vary across the groups and be representative of a larger population of intercepts associated with a population of groups from which the 50 groups in the study were selected. The mean $\overline{X}_g$ is the mean for X (i.e., pay) within group g. Substituting Equation 22 into Equation 21, we obtain the combined model

$$Y_{ig} = \gamma_0 + \gamma_1 X_{ig} + \gamma_2 \overline{X}_g + u_{0g} + r_{ig}, \tag{23}$$

and the coefficient γ_2 in Equation 23 represents the context or compositional effect as it provides information about how group composition predicts the dependent variable.

The second alternative in MLM involves substituting the term $(X_{ig} - \overline{X})$, the deviation of pay for employee i in group g from mean pay for all employees, in place of X_{ig} in the Level 1 model (Equation 21), with all other aspects of the modeling in Equations 21 and 22 remaining the same. Thus, the combined model for this grand-mean centering approach (see Raudenbush & Bryk, 2002) is

$$Y_{ig} = \gamma_0 + \gamma_1 (X_{ig} - \overline{X}) + \gamma_2 \overline{X}_g + u_{0g} + r_{ig}, \tag{24}$$

where, again, the coefficient γ_2 represents the context or compositional effect.

The combined model for the final alternative, or group-mean centering approach, is

$$Y_{ig} = \gamma_0 + \gamma_1(X_{ig} - \overline{X}_g) + \gamma_2^* \overline{X}_g + u_{0g} + r_{ig} \tag{25}$$

and involves substituting the term $X_{ig} - X_g$, or the deviation of pay for employee i in group g from mean pay for group g, in place of X_{ig} in the Level 1 model (Equation 21). As Equation 25 also shows, however, the coefficient of $\overline{X}_g$ is γ_2^*, as this effect in Equation 25 does not equal that of γ_2 as estimated under the other two alternatives (Equations 23 and 24). The coefficient γ_2^* is called the between-groups coefficient, and it can be shown that $\gamma_2 = \gamma_2^* - \gamma_1$. That is, if the between-groups coefficient (γ_2^*) and the within-groups coefficient (γ_1) are not equal, there is context effect. Moreover, γ_1 can simply be estimated by using

$$Y_{ig} = \gamma_0 + \gamma_1(X_{ig} - \overline{X}_g) + u_{0g} + r_{ig}, \tag{25a}$$

which results in the same coefficient (γ_1) for the continuous independent variable as is obtained by including $\overline{X}_g$ as a separate independent variable when using the group-mean centering approach. Using this latter equation to estimate the relationship of job satisfaction to individual pay in the Klein et al. (2000) data, $\hat{\gamma}_1 = .97$ and the F statistic for testing the null hypothesis $\gamma_1 = 0$ is $F(1, 699) = 151.56$, $p < .05$.

It is important to note that for all of the alternative MLM approaches, the population coefficient γ_1 is the same, estimates of γ_1 are the same, and tests on γ_1 are the same as long as $\overline{X}_g$ is included in the model when the raw form (Equation 23) or grand-mean-centered form (Equation 24) of the independent variable are used. Although $\overline{X}_g$ is incorporated in the group-mean centering approach (Equation 25a) in the independent variable $X_{ig} - \overline{X}_g$, failure to include $\overline{X}_g$ as a separate independent variable in the raw form (Equation 23) and grand-mean-centered (Equation 24) approaches inherently assumes that the group-mean score on the independent variable (organization-specific mean salary in our example) does not provide any predictive power over and above the individual score on the independent variable (individual salary in our example). If this assumption is not correct, however, leaving out $\overline{X}_g$ as an independent variable in these latter two approaches results in a misspecified model. Raudenbush and Bryk (2002) argued that when $\overline{X}_g$ is omitted when it should have been included as an independent variable, the resulting estimate of γ_1 will likely be misleading because it will be an uninterpretable weighted average of $\hat{\gamma}_1$ (the within-groups coefficient) and $\hat{\gamma}_2^*$ (the between-groups coefficient; see Raudenbush & Bryk, 2002, pp. 134-139, for a more complete discussion). In short, when investigating the relationship of a continuous dependent variable to a continuous within-groups independent variable in multilevel data, $\overline{X}_g$ should be included as an independent variable.

RMR can also be used to investigate the within-groups relationship of job satisfaction to pay. The formula for the F statistic for testing the null hypothesis $\gamma_1 = 0$ is

$$F = (N - G - 1)\frac{R^2_{inc,Pay}}{(1 - R^2_{Bet}) - R^2_{inc,Pay}}. \tag{26}$$

In Equation 26, R^2_{Bet} is the same as above (Equations 11 and 12). The quantity $R^2_{inc,Pay}$ can be computed based on the regression model

$$Y_{ig} = \beta_0 + \beta_1 X_{ig} + \beta_2 \overline{X}_g + \varepsilon_{ig} \tag{27}$$

and is the increase in the R^2 for this model when X_{ig}, individual pay, is added to the equation containing $\overline{X}_g$, group mean pay. For Equation 27, $\beta_1 = \gamma_1$, and for the Klein et al. (2000) data, $\hat{\beta}_1 = 0.97$. Substitution in the formula for F (Equation 26) yields

$$F = (750 - 50 - 1)\frac{.14836}{(1 - .16744) - .14836} = 151.56$$

and has 1 and 699 degrees of freedom, which is the same result as found using MLM.

Although MLM and RMR yield the same F statistics in a balanced design under this set of assumptions (i.e., that the relationship between pay and job satisfaction is the same for all groups), in an unbalanced design, there will be some difference between the two F statistics. However, in the cases we have investigated, the difference in the MLM and RMR F statistics is very small. We again randomly deleted cases from the Klein et al. (2000) data with a .5 probability of deletion. For MLM, $\hat{\gamma}_1 = .95$ and $F(1, 331) = 65.50, p < .0001$; for RMR, $\hat{\beta}_1 = .95$ and $F(1, 331) = 65.21, p < .0001$.

The relationship of job satisfaction to pay varies across groups. More substantial differences between the two approaches emerge when we abandon the assumption that the within-group regression coefficient is the same for all groups and instead assume that it varies randomly across the groups. The consideration of such a model is highly relevant given that in the context of a group, characteristics of that group can influence relationships between individual-level variables. In fact, this type of cross-level interaction effect (Rousseau, 1985) is often the underlying focus of research taking place in the context of groups and teams (e.g., Klein et al., 2000; LePine & Van Dyne, 1998).

In MLM, relaxing the assumption that within-group regression coefficients are the same for all groups results in the following Level 1, Level 2, and combined models (using the raw form approach):

$$\text{Level 1 model: } Y_{ig} = \beta_{0g} + \beta_{1g}X_{ig} + r_{ig} \tag{28}$$

$$\text{Level 2 model: } \beta_{0g} = \gamma_0 + \gamma_2\overline{X}_g + u_{0g} \tag{29}$$

$$\beta_{1g} = \gamma_1 + u_{1g} \tag{30}$$

$$\text{combined model: } Y_{ig} = \gamma_0 + \gamma_1 X_{ig} + \gamma_2\overline{X}_g + u_{0g} + u_{1g}X_{ig} + r_{ig}. \tag{31}$$

This combined model is called an intercepts-as-outcomes model. It specifies that the groups are selected from a larger population of groups, and the intercept and slopes vary over the groups and thus can be modeled. For the balanced Klein et al. (2000) data, $\hat{\gamma}_1 = .98$, and the F statistic for testing the null hypothesis $\gamma_1 = 0$ is $F(1, 49) = 114.76, p < .05$.

The RMR approach, however, cannot handle this relaxation of assumptions. The conventional method of using RMR, as represented in Equations 26 and 27, does not allow for effects to vary randomly across groups. Given the F statistic produced by this approach, $F(1, 699) = 151.57$, it is clear that even in a balanced design, the RMR approach can give results dramatically different from those obtained with the MLM random coefficients approach.

Summary. In summary, for multilevel designs in which the focus is assessing the effects of a continuous individual-level independent variable, we suggest the following:

1. When the effect of the individual-level independent variable is assumed to be the same across all groups, the proper methodology to use depends on whether the design is balanced. For balanced designs, RMR and MLM will result in the same F values as long as the researcher specifies a random intercept model when using MLM and that the context effect of $\overline{X}_g$ is included in the estimation, either via group-mean centering or as a separate independent variable when using raw form or grand-mean centering. For unbalanced designs, although RMR and MLM results for a random intercept model will likely be similar to one another, we recommend using MLM because the operating characteristics of RMR are unknown. As with the last design, the same caveat applies regarding small samples.
2. When the effect of the individual-level independent variable is assumed to randomly vary across groups, the MLM random coefficients approach should be used. There is no direct way to assess a random coefficient model with RMR, and F tests for possible regression-based approaches differ a great deal from those obtained using MLM.

Discussion

Scholars are increasingly interested in questions that are inherently multilevel in nature. Unfortunately, however, the literature is not clear regarding which type of analytic approach to use in different types of multilevel research designs. In this article, we build on the work of Klein and colleagues (2000) and compare the use of RMR with repeated-measures ANOVA, multivariate repeated-measures ANOVA, and multilevel modeling in four different and popular multilevel research designs. We focus on the suitability of RMR for multilevel research for several reasons. First, although it has been adopted in certain segments of the literature to assess multilevel questions, its capabilities for such purposes have been relatively unexplored compared to such techniques as MLM. Second, RMR is relatively easy to learn by those who are familiar with ANOVA and hierarchical regression, and it also explicitly models within and between variance similar to sophisticated MLM techniques that might require a great deal of learning and thus is a potentially appealing technique to researchers.

Overview of Results

Table 7 summarizes our findings in four different sections representing each research design examined. Each section contains the questions/assumptions relevant to the design and then lists all those methods that may be viable under each scenario, based on the adequacy of their parameter estimates and statistical tests. In the cases in which more than one method may be adequate, we suggest that the choice of method should then be driven by what the particular researcher is most comfortable using (i.e., ease of use) and also consideration of consistency in reporting and presenting the results. That is, we suspect that readers will likely find it easier to follow results when the same analytic method is consistently used throughout the entire article if this is possible.

As Table 7 shows, we found that in general, sophisticated MLM approaches are not always necessary to analyze multilevel data. In the case of a simple repeated-measures design, for example, ANOVA appears to be adequate and MLM might be necessary only if there are missing data or if other assumptions besides the sphericity and unstructured assumption are

Table 7
Summary of Recommendations

Multilevel Design	Relevant Questions/Assumptions and Recommendations			
		Method to Use With		
	Sphericity Assumption Met?	Balanced Data	Missing Data	Alternative Assumptions of Variance/Covariance
Designs with one within-groups factor	Yes	ANOVA, MANOVA, RMR, MLM	MLM	MLM
	No	ANOVA, MANOVA, MLM	MLM	MLM
	No	ANOVA, MANOVA, MLM		
		Balanced Data?	Method to Use When Estimating σ_B^2 and σ_W^2	Method to Use When Testing the Hypothesis $\sigma_B^2 = 0$
Designs with group membership as the only between-groups factor		Yes	ANOVA, MLM	ANOVA
		No	MLM	ANOVA
		Balanced Data?	Method to Use	
Designs with one continuous group-level independent variable		Yes	RMR, MLM	
		No	MLM	
		Assumption Regarding the Slope of the Relationship	Balanced Data?	Method to Use
Designs with one continuous individual-level independent variable		Same across groups	Yes	RMR, MLM
			No	MLM
		Varies across groups	Yes	MLM
			No	MLM

Note: ANOVA = repeated-measures analysis of variance; MANOVA = multivariate repeated-measures analysis of variance; RMR = repeated-measures regression; MLM = multilevel modeling.

made with regard to the variance/covariance matrix (see first section of Table 7). It should be noted, however, that this recommendation presumes that researchers using ANOVA for such designs correct for assumption violations (i.e., sphericity) when appropriate. Considering that Bergh (1995), who performed a content analysis of repeated-measures studies appearing in seven management journals for the period from 1985 to 1992, found that only 3.4% of the studies controlled for assumption violations, the use of ANOVA for such designs still may be problematic. If researchers are not going to properly correct for assumption violations, then multivariate repeated-measures ANOVA or MLM approaches may be preferable. The second section of Table 7 also shows that when it comes to estimating the proportions of between- and within-group variance, the use of MLM would seem to be necessary only with unbalanced data; ANOVA is more than adequate with balanced data, given its ease of use. Furthermore, ANOVA is the preferred method for testing the hypothesis $\sigma_B^2 = 0$. It is clear, however, that increased sophistication in analytic techniques becomes more necessary as the research design and question become more complex. As the third and fourth sections of Table 7 show, when designs incorporate a continuous independent variable at either the group or individual level, although RMR may be used with balanced data, in general, MLM becomes mandatory.

At a more specific level with respect to RMR, our assessment is that the technique may be adequate in a relatively small number of situations. As the first section of Table 7 shows, in a simple repeated-measures design, RMR may be used as long as there are no missing data and as long as the researcher is confident that the sphericity assumption is not violated. However, for this design, there is no advantage to RMR over the other techniques. In a simple between-groups design in which the only information is group membership (second section, Table 7), RMR is not even listed as a viable method because, although the F test will be the same as ANOVA in a balanced design, the R^2 does not provide the correct estimate of the proportion of variance attributed to be between groups. For a design in which there is a single continuous group-level variable (third section, Table 7), RMR will produce correct coefficient estimates and F tests as long as the design is balanced. For a design in which there is a continuous individual-level variable (fourth section, Table 7), RMR may be used as long as the design is balanced and the assumption that the slopes do not vary across groups is reasonable. When this assumption is not reasonable, RMR is no longer appropriate, however, because the F test it produces is incorrect when the slopes vary across groups. Therefore, MLM becomes necessary.

This last finding suggests that although versions of models for investigating more complex research designs can be formulated using variations on RMR, their utility for more complex research designs is severely limited. For example, examinations of cross-level interactions, such as the moderating effect that a group-level variable may have on the effect of an individual-level variable to some outcome, are quite common in organizational research. In the context of the Klein et al. (2000) data, we might ask whether mean pay predicts the slope of the individual-level relationship between job satisfaction and pay. Perhaps if a group's average pay is sufficiently high, individual-level pay would not be related to job satisfaction. RMR is able to examine such a question in only a very limited way. Although RMR may be able to examine the interaction between group mean pay and individual-level pay on individual job satisfaction, RMR is unable to assess the random coefficient models that are necessary to generalize its findings regarding cross-level interactions beyond the sample of groups (organization, etc.) included in the study. In other words, because the method may be used only under the assumption that slopes are fixed across groups, any findings of such cross-level interactions in RMR are applicable only to the groups under study—no inference may be made to some larger population of interest. Such inference can be made, however, using the random coefficients modeling of MLM.

Another serious shortcoming of RMR is that its practical utility, at least as compared to MLM, becomes questionable as the number of variables included in the model increase. The convention followed by previous studies using RMR has involved a Type I sums of squares approach—the adjusted F ratio is based on changes in R^2 associated with each effect, accounting only for the effects that precede it in the model—and has involved entering the individual-level effects before the group-level variables (e.g., Hollenbeck et al., 1994). Thus, the adjusted F ratios in this conventional RMR approach do not take into account effects from the other levels of analysis (i.e., within-group F ratio does not take into account group-level effects). Although this convention is rather straightforward and is not a problem in the case in which all effects in the specified model are uncorrelated, it falls prey to misspecification or omitted variable bias when effects are correlated, especially between individual- and group-level variables. Therefore, a Type III sums of squares approach (which includes variation that is unique to an effect after adjusting for all other effects that are included in the model, regardless of the level of analysis) is necessary when the effects in the model are correlated. As we demonstrated in the last section of this study (i.e., Equations 26 and 27), the Type III sums of squares approach is possible in RMR. The use of RMR in such a manner would become quite cumbersome, nevertheless, as the number of variables under study increase: As many separate estimations would have to be run as there are variables (with each variable entered into the modeling last) and the adjusted F ratios subsequently all calculated manually. In contrast, the Type III sums of squares approach is the default method of significance testing of effects in MLM and thus produces results for all effects in the modeling simultaneously and automatically.

Ultimately, if RMR is to become a useful technique, additional research aimed at developing a better understanding of the operating characteristics of RMR (i.e., Type I and Type II error rates) given different types of data and research designs (e.g., unbalanced design) is needed. Knowing whether RMR can be used for more complex research designs requires demonstrating that it provides the same results as the correct analytic approach, and as the nature of the multilevel design becomes more complex, MLM appears to become that correct approach. Thus, given the limited utility of RMR for simple research designs demonstrated in the current research, its obvious limitations for more complex research, and the existence of better suited multilevel techniques, such additional research on RMR is perhaps futile.

Limitations

The scope of this article was limited in at least two ways. First, our research may appear to be focused on a relatively narrow analytic topic. However, researchers are increasingly considering questions that are inherently multilevel, and it is critical that the field develop an appreciation of the relevant analytic issues and also some knowledge about how to deal with these issues. We illustrate that if the wrong multilevel analytic technique is used, parameter estimates and tests of statistical significance may be incorrect. Given that organizationally relevant decisions are made based on these estimates and tests, it is critical that we understand which technique to use in a given situation.

Second, we did not consider several potentially useful multilevel analytic techniques. For example, there are relatively simple to use statistical software programs that take into account nesting of data when assessing individual-level relationships. The software program SUDAAN (Shah, Barnwell, & Bieler, 1997), for example, uses generalized estimating equations (Zeger & Laing, 1986; Zeger, Liang, & Albert, 1988) to calculate standard errors for regression weights that take into account intracluster/intragroup correlations. SAS has this capability as well. LePine, Colquitt, and Erez (2000), for instance, used SUDAAN to calculate standard errors in a repeated-measures design in which the research question focused on predicting an individual's adaptation to a changing task context.

Multiple indicator latent growth modeling is another example of a more complex multilevel analytic technique that we did not consider in this article. In fact, this approach may be preferable to MLM (as well as the other approaches that we consider) when the researcher is interested in modeling intraindividual change (Chan, 1998a). For example, scholars have used the technique to understand factors that influence newcomer adaptation and social development (e.g., Chan, Ramey, Ramey, & Schmitt, 2000; Chan & Schmitt, 2000; Schmitt, Sacco, Ramey, Ramey, & Chan, 1999). Relative to MLM, one advantage of latent growth modeling is that it is able to consider measurement error. This capability of latent growth modeling makes it possible to specify and assess alternative time-related error covariance structures, and as a result, model misspecification that results in biased estimates of the change and change patterns may be avoided (Chan, 1998a). Although it may have been worthwhile to consider latent growth modeling in our research, we did not do so because the most relevant comparison with this technique would have been with MLM, and our research was focused more on comparisons with RMR. Moreover, Chan (1998a) provided a very comprehensive overview of latent growth modeling that included comparisons with MLM.

Conclusion

For relatively simple designs, ANOVA or multivariate repeated-measures ANOVA were generally as good as or superior to RMR. For more complex designs, MLM is generally as good as or superior to RMR. Moreover, using RMR to obtain the correct results in relatively complex multilevel research designs is not always a straightforward or even possible task. Thus, although RMR will produce correct results in some circumstances, it seems to us that researchers would be well served to learn the ins and outs of MLM. Indeed, the new addition of the classic Cohen and Cohen text (Cohen, Cohen, West, & Aiken, 2000) includes a very good treatment of MLM.

References

Algina, J., & Keselman, H. J. (1997). Detecting repeated measures effects with univariate and multivariate statistics. *Psychological Methods*, *2*, 208-218.

Beersma, B., Hollenbeck, J. R., Humphrey, S. E., Moon, H., Conlon, D. E., & Ilgen, D. R. (2003). Cooperation, competition, and team performance: Toward a contingency approach. *Academy of Management Journal*, *46*, 572-590.

Bergh, D. D. (1995). Problems with repeated measures analysis: Demonstration with a study of the diversification and performance relationship. *Academy of Management Journal*, *38*, 1692-1708.

Berkhof, J., & Snidgers, T. A. B. (2001). Variance component testing in multi-level models. *Journal of Educational and Behavioral Statistics*, *26*, 133-152.

Boik, R. J. (1981). A priori tests in repeated measures designs: Effects of nonsphericity. *Psychometrika, 46*, 241-255.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the analysis of variance: II. Effects of inequality of variance and correlation between errors in the two-way classification. *Annals of Mathematical Statistics, 25*, 484-498.

Chan, D. (1998a). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods, 1*, 421-483.

Chan, D. (1998b). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*, 234-246.

Chan, D. (2002). Longitudinal modeling. In S. G. Roselberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 412-430). Malden, MA: Blackwell.

Chan, D., Ramey, S. Ramey, C., & Schmitt, N. (2000). Modeling intraindividual changes in children's social skills at home and at school: A multivariate latent growth approach to understanding between-settings differences in children's social skill development. *Multivariate Behavioral Research, 35*, 365-396.

Chan, D., & Schmitt, N. (2000). Interindividual differences in intraindivual changes in proactivity during organizational entry: A latent growth modeling approach to understanding newcomer adaptation. *Journal of Applied Psychology, 85*, 190-210.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlational analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Gully, S. M. (1994, April). *Repeated measures regression analyses: A clarification with illustrative examples.* Paper presented at the ninth annual conference of the Society for Industrial and Organizational Psychology, Nashville, TN.

Gully, S. M. (1997, April). *Effect size, theory, and tests of significance in repeated measures and multi-level analyses.* Paper presented at the 12th annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Hofmann, D. A., Griffin, M. A., & Gavin, M. B. (2000). The application of hierarchical linear modeling to organizational research. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multi-level theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 467-511). San Francisco: Jossey-Bass.

Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., Colquitt, J. A., & Hedlund, J. (1998). Extending the multi-level theory of team decision making: Effects of feedback and experience in hierarchical teams. *Academy of Management Journal, 41*, 269-282.

Hollenbeck, J. R., Ilgen, D. R., & Sego, D. J. (1994). Repeated measures regression and mediation tests: Enhancing the power of leadership research. *Leadership Quarterly, 5*, 3-23.

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multi-level theory of team decision making: Decision making in teams incorporating distributed expertise. *Journal of Applied Psychology, 80*, 292-316.

Hox, J. (2002). *Multilevel analysis techniques and applications.* Hillsdale, NJ: Lawrence Erlbaum.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics, 1*, 69-82.

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology, 54*, 1-20.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 52*, 63-78.

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. (1980). Testing the validity of conditions of repeated measures *F* tests. *Psychological Bulletin, 87*, 479-481.

Klein, K. J., Bliese, P. D., Kozlowski, S. W. J., Dansereau, F., Gavin, M. D., Griffin, M. A., et al. (2000). Multilevel analytical techniques: Commonalities, differences, and continuing questions. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multi-level theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 512-553). San Francisco: Jossey-Bass.

Klein, K. J., & Kozlowski, S. W. J. (Eds.). (2000). *Multi-level theory, research, and methods in organizations: Foundations, extensions, and new directions.* San Francisco: Jossey-Bass.

Klein, K. J., Tosi, H., & Cannella, A., Jr. (1999). Multi-level theory building: Benefits, barriers, and new developments. *Academy of Management Review, 24*, 243-248.

LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology, 53*, 563-593.

LePine, J. A., & Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of Applied Psychology, 83*, 853-868.

Lorch, R. F., Jr., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology, 16*, 149-157.

Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology, 85*, 971-986.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: John Wiley.

Padilla, M. A., & Algina, J. (2004). Type I error rates for a one factor within-subjects design with missing values. *Journal of Modern Applied Statistical Methods, 3*, 406-416.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). Fort Worth, TX: Holt, Rinehart & Winston.

Phillips, J. M. (2001). The role of decision influence and team performance in member self-efficacy, withdrawal, satisfaction with the leader, and willingness to return. *Organizational Behavior and Human Decision Processes, 84*, 122-147.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models applications and data analysis methods*. Thousand Oaks, CA: Sage.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross level perspectives. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 1-38). Greenwich, CT: JAI.

Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359-391.

Schmitt, N., Sacco, J. M., Ramey, S., Ramey, C., & Chan, D. (1999). Parental employment, school climate, and children's academic and social development. *Journal of Applied Psychology, 84*, 737-753.

Shah, B. V., Barnwell, B. G., & Bieler, G. S. (1997). *SUDAAN user's manual, release 7.5*. Research Triangle Park, NC: Research Triangle Institute.

Wolfinger, R. D., & Chang, M. (1995). Comparing the SAS GLM and MIXED procedures for repeated measures. *SAS User's Group International Conference Proceeding SUGI, 20*, 1172-1182.

Zeger, S., & Liang, K. (1986). Longitudinal data analysis for discrete and continuous outcomes, *Biometrics, 42*, 121-130.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics, 44*, 1049-1060.

Vilmos F. Misangyi is an assistant professor of management in the Alfred Lerner College of Business & Economics at the University of Delaware. He received his Ph.D. in strategic management from the University of Florida. His research focuses on CEO charisma, corporate governance, and multilevel analysis.

Jeffery A. LePine is an associate professor of management in the Warrington College of Business at the University of Florida. He received his Ph.D. in organizational behavior from Michigan State University. His research has focused on topics associated with team adaptability and effectiveness, citizenship performance, occupational stress, and individual differences.

James Algina is a professor of educational psychology at the University of Florida. His research interests are in applied statistics and psychometrics.

Frank Goeddeke Jr. analyzes the domestic automobile industry in the National Organizing Department of the United Auto Workers Union, Detroit, Michigan. He received his Ph.D. in management from the University of Florida and his M.B.A. from the Crummer Graduate School of Business, Rollins College, Winter Park, Florida.