2007

Multilevel Models for Examining Individual Differences in Within-Person Variation and Covariation Over Time

Lesa Hoffman

University of Nebraska-Lincoln, lhoffman2@unl.edu

Multilevel Models for Examining Individual Differences in Within-Person Variation and Covariation Over Time

Lesa Hoffman
University of Nebraska-Lincoln

Heterogeneity of variance may be more than a statistical nuisance—it may be of direct interest as a result of individual differences. In studies of short-term fluctuation, individual differences may relate to the magnitude of within-person variation as well as to level of an outcome or its covariation with other processes. Although models for heterogeneous variances have been utilized in group contexts (i.e., dispersion models), they are not usually applied in examinations of intraindividual variation. This work illustrates how an extension of the multilevel model for heterogeneous variances can be used to examine individual differences in level, between- and within-person covariation, and magnitude of within-person variation of daily positive and negative mood in persons with dementia.

Hypotheses about psychological and developmental processes are frequently, if not almost always, centered on the detection of differences in level between individuals or between groups on a given outcome. For example, the effect of a continuous predictor on an outcome is usually evaluated as the strength of the linear relationship between level of that predictor and level of the outcome. Similarly, the efficacy of an intervention or experimental manipulation is usually assessed as the extent to which mean differences are found between treatment and control groups or between design conditions. Within-group variation in such cases is usually regarded as a statistical nuisance—as the noise from which the signal of an effect must be separated. Accordingly, analytic methods for addressing violations of homogeneity of variance have focused primarily on

Correspondence concerning this article should be addressed to Lesa Hoffman, Department of Psychology, University of Nebraska-Lincoln, 238 Burnett Hall, Lincoln, NE 68588-0308. E-mail: lhoffman2@unlnotes.unl.edu

obtaining correct inferences about mean differences in such cases (e.g., Aguinis & Pierce, 1998; DeShon & Alexander, 1996; Grissom, 2000; Overton, 2001).

Differential within-group variation need not be regarded as merely as a nuisance. Psychological or developmental processes might exert their effects not only on the level of an outcome but also on variation of that outcome within groups or within persons (Bryk & Raudenbush, 1988; Olejnik, 1998). The extent to which persons differ in within-task or across-task variability and the extent to which variability is related across domains has been of considerable recent interest in many areas and within the study of cognitive aging in particular. Cross-sectional studies have reported greater within-person, within-task variability in cognitive tasks in older persons (Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000), as well as greater within-person, across-task variability with age in nursing home residents (Rapp, Schnaider-Beeri, Sano, Silverman, & Haroutunian, 2005). Longitudinal studies have also suggested that greater within-person variability or inconsistency is predictive of greater cognitive decline (Kliegel & Sliwinski, 2004; MacDonald, Hultsch, & Dixon, 2003) and is associated with lower intelligence (Ram, Rabbitt, Stollery, & Nesselroade, 2005). Finally, within-person variability has also been found to relate across domains of cognition and physical function in persons with dementia, suggesting that magnitude of variability may serve as a marker of neurological integrity (Strauss, MacDonald, Hunter, Moll, & Hultsch, 2002).

In the aforementioned studies and in others, within-person variation is characterized as a trait of an individual. Accordingly, an individual-level summary measure of that variation is first computed, often as a within-person standard deviation (SD). These summary measures are then used in subsequent between-person analyses, such as examining predictors of between-person differences in the magnitude of within-person SDs, or the extent to which within-person SDs are correlated across domains. An alternative to examining covariation as an individual-level trait is a *within-person* approach, which provides a more direct basis for making inferences at the intraindividual level (Sliwinski & Buschke, 2004). A within-person analysis addresses the extent to which domains covary over time within an individual or what time-varying covariates predict within-person variation (i.e., within-person covariation). That is, on a given occasion, if a person scores high on one domain, relative to his or her usual level, does that person also score high on another domain, again relative to his or her usual level? Between-person analyses can also be used to examine individual differences in heterogeneity of within-person variation or individual differences in within-person covariation (Baltes & Nesselroade, 1979).

The purpose of this article is to illustrate a general analytic framework for addressing such between-person and within-person questions of variation and covariation simultaneously. The multilevel or general linear mixed model is a well-known tool for examining individual differences in between- and within-

person covariation of time-dependent processes. In addition, however, rather than using a two-stage approach of computing within-person *SD*s to examine differential *magnitude* of within-person variation, a more parsimonious and potentially more powerful approach lies within an extension of the multilevel model for heterogeneous variances. Such models have been referred to as *dispersion models* in clustered or nested data (e.g., students nested in classes), such as in the educational literature, in which differences between classes in within-class heterogeneity is often of importance. Raudenbush and Bryk (1987) presented a framework for estimating such models as multilevel models using maximum likelihood and provided an example in which within-school dispersion or heterogeneity in math achievement was found to relate to school-level predictors. Yet despite their direct applicability, multilevel models with heterogeneous variances largely have not been used in examining individual differences in short-term within-person variation, a context for which they are well suited, perhaps in part because published examples of these models deal primarily with the nested case of within-group heterogeneity (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). This work aims to fill this gap by illustrating these models in a within-person context.

AN ILLUSTRATIVE EXAMPLE

The data for this example were collected during the baseline phase of a clinical trial from a sample of 31 nursing home residents with dementia. Self-reported positive and negative mood were collected daily for 12 days. The first goal of the analysis was to examine stability of self-reported mood across days and individual differences in that stability. It was expected that positive and negative mood will be less stable (i.e., show greater within-person variation) in persons with lower mental functioning, given that external or contextual factors are thought to have a greater influence on mood in such persons (Smith, Gerdner, Hall, & Buckwalter, 2004).

The second goal of the analysis was to examine covariation in positive and negative mood. The reliability of the information about emotional well-being provided by persons with dementia is often of considerable debate. One way in which reliability can be assessed is through internal consistency. If the respondents understood the mood questions and answered in a way that accurately reflected their mood on a given day, then internal consistency should manifest itself as an inverse relationship between positive and negative mood. However, an analysis of the relationship between positive and negative mood must take into account the fact that mood varies both between-persons as well as within-persons. For instance, some people are in better moods than other people on average (between-person variation), and people are in better moods than usual

on some days than other days (within-person variation). Thus, covariation between positive and negative mood at both levels should be examined to see if respondents rated themselves consistently relative to their peers and relative to their usual levels of mood. Finally, if covariation is indeed an index of internal consistency, one might expect that consistency to be greater in individuals with greater mental functioning. Thus, a third goal of the analysis was to examine moderation of between- and within-person covariation in mood by mental functioning, as measured by the Mini-Mental State Exam (MMSE). If internal consistency is indeed greater in persons with greater mental functioning, then negative covariation should be stronger (i.e., more negative) at greater levels of MMSE.

In summary, three questions directed the analyses: (1) Do level and stability over 12 days of positive and negative mood differ by mental functioning? (2) Do positive and negative mood covary between- and within-persons? and (3) Does between- and within-person covariation of positive and negative mood (internal consistency) differ by mental functioning?

METHOD

Participants and Design

The study sample included 31 nursing home residents (24 women) with dementia who were recruited from four nursing homes in central and northeast Pennsylvania and who met strict enrollment criteria (see Kolanowski, Litaker, & Buettner, 2005, for more information). On average, the residents were 82.7 years old ($SD = 7.7$, range $= 58$ to 94) with 11.0 years of education ($SD = 2.5$, range $= 6$ to 16). This example used baseline data over 12 days from a crossover experimental study that tested the efficacy of three treatments for reducing agitation and passivity in nursing home residents with dementia. Participants were observed and videotaped for 20 min each day at the time of day when they exhibited a high level of agitation or passivity as determined by staff report and observation. Prior to and at the completion of each 20-min observation session, participants were asked about their mood using a standard instrument (described later) by a trained research assistant blind to study aims. Only the measures taken at the beginning of the 20-min period were used for analysis. Further information about the study is also available in Kolanowski, Hoffman, and Hofer (2007).

Measures

Mental functioning.   Mental functioning was measured by the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975). The MMSE items

assess orientation, registration, attention, calculation, recall, language, and visual construction, and each item has between two and five categories. The score is the sum of all the correct answers (range of 0–30), with higher scores indicating greater mental functioning. The mean MMSE for the sample was 8.61 ($SD =$ 7.14, range $= 0$ to 26), indicating moderate to severe impairments.

Self-reported mood.   Mood was measured in real time using the Dementia Mood Picture Test (Tappen & Barry, 1995), an instrument that measures both positive and negative moods from the perspective of the cognitively impaired participant. The participant was shown pictures of six faces and asked to indicate if the drawing represented how he or she felt at that time. The six faces were designed to portray *bad mood*, *good mood*, *angry*, *sad*, *happy*, and *worried*, with possible response options to each of *no*, *yes*, and *very much*.

Because the item responses were non-interval (i.e., the difference in mood between answering *no* and *yes* is not likely to equal the difference in mood between answering *yes* and *very much*), a simple sum score of the item responses as 0, 1, or 2 may be misleading. Instead, a two-factor graded response Rasch model was estimated in Mplus 3.13 (Muthén & Muthén, 1998–2004) to obtain latent traits for positive mood (as indicated by pictures of *good mood* and *happy*) and negative mood (as indicated by pictures of *bad mood*, *angry*, *sad*, and *worried*). The latent trait estimates were then used in subsequent multilevel analyses.

Due to the small sample size ($n = 31$), the model was estimated on the combined person-period data set (i.e., 31 persons by 12 days). In doing so, however, there are at least three problems to be acknowledged. The first is that the latent trait estimates will not have the same properties as the latent variable from which they were derived (Grice, 2001; Lu, Thomas, & Zumbo, 2005). The second is that one must assume measurement invariance over time and equivalence of the between- and within-person measurement model because the sample size is prohibitive in testing these assumptions. Given that the data were collected over 12 exchangeable days, however, the assumption of invariance across time is likely to be reasonable. Further, invariance over time and equivalence of the between- and within-person measurement model of the outcomes are assumed in most longitudinal analyses based on observed variables, whether explicitly acknowledged or not. The third problem is that the analysis ignores the dependency in the data (i.e., that residuals from the same person are likely to be correlated). Although such dependency is known to impact the standard errors of estimated coefficients, it is not likely to bias the estimates themselves (see Cohen, Cohen, West, & Aiken, 2003). It is the latent trait estimates (and not their standard errors) that will be used in a subsequent multilevel analysis in which the dependency of observations from the same person will indeed be modeled properly. Finally, although analytic methods for single-stage

estimation are undergoing development, extensions for the simultaneous modeling of heterogeneous variances are not currently available. For these reasons, a two-stage procedure was preferred in order to account for the non-equal interval between the item response options and to eliminate measurement error to the greatest extent possible in the daily mood outcomes to be modeled (see also Curran, Edwards, Wirth, & Hussong, 2007, and Osgood, McMorris, & Potenza, 2002).

RESULTS AND DISCUSSION

Between-person and within-person variation in self-reported mood was examined using multilevel models (Littell, Milliken, Stroup, & Wolfinger, 1996; Snijders & Bosker, 1999) estimated in SAS PROC MIXED. (The electronic appendix containing the data and SAS syntax used for analysis is available at http://psycweb.unl.edu/psypage/hoffman/HomePage.htm)

Models differing in fixed effects were compared using maximum likelihood (ML), and models differing in error structure or random effects only were compared using restricted maximum likelihood (REML). Nested models were compared by their model deviances (-2 log likelihood values) as a function of the difference in the number of parameters estimated in each, and non-nested models were compared by information criteria. The significance of fixed effects was evaluated with Wald's tests with Satterthwaite denominator degrees of freedom. The Satterthwaite method is recommended in smaller sample sizes for which a t of F distribution is preferred over a standard normal distribution but in which the data are unbalanced and thus the calculation of denominator degrees of freedom is not straightforward (see Fitzmaurice, Laird, & Ware, 2004). An unstructured matrix was estimated for any random effects (i.e., all random variances and covariances in the G matrix estimated separately). The specific models to be estimated are presented next, and the sequence of decisions that follows is summarized in Figure 1.

Unconditional Models of Stability in Positive and Negative Mood

Random intercept model.   One way of examining stability across the 12 days in positive and negative mood is by estimating an unconditional (i.e., without predictors) random intercept two-level model, as shown in Equation 1:

$$\text{Level 1:} \quad y_{di} = \beta_{0i} + e_{di}$$

$$\text{Level 2:} \quad \beta_{0i} = \gamma_{00} + U_{0i} \tag{1}$$

1. Decomposition of variation with an empty model.

Report the intraclass correlation.

2. Treatment of time with respect to the means:
Is there a systematic effect of time on the means that should be represented (e.g., day-of-week or day-of-study effects)?

YES. Include fixed effects for time in the model for the means. Keep in mind that the residual variance will reflect mis-fit of the model for time as well as within-person variation.

NO. Include a fixed intercept only in the model for the means.

3a. Treatment of time with respect to the variances:
Are there individual differences in the systematic effects of time (do the effects of time need to be random?)

YES. Include random effects for time (and their covariances) in the model for the variances.

NO. Include a random intercept only in the model for the variances.

3b. Treatment of time with respect to the variances:
Are the residuals uncorrelated across time? Does a correlated residual model fit better than an uncorrelated residual model?

YES. Include parameter(s) for the correlation in the matrix of residuals (e.g., AR1, Toeplitz). This can be included with or without the random intercept.

NO. Use an identity structure for the matrix of residuals (the default).

4. Test for homogeneity of residual variance:
Is the test for homogeneity of variance non-significant?
Do plots of the residuals appear homogeneous across persons?

YES. There is little evidence for individual differences in within-person variation, so predictors of the residual variance may not be useful or even necessary.

NO. Examine predictors of individual differences in within-person variation (included as fixed effects, too).

5. Incorporate between-person and within-person predictors of interest as fixed or random effects.

6. Evaluate alternative explanations for heterogeneity:
Is the Level-1 model correctly specified?
Are the residuals normally distributed?

YES. Heterogeneity of within-person variation is not likely to be the result of improper model specification or violation of model assumptions.

NO. Consider models with alternative predictors. Also consider transformations to improve normality of the residuals.
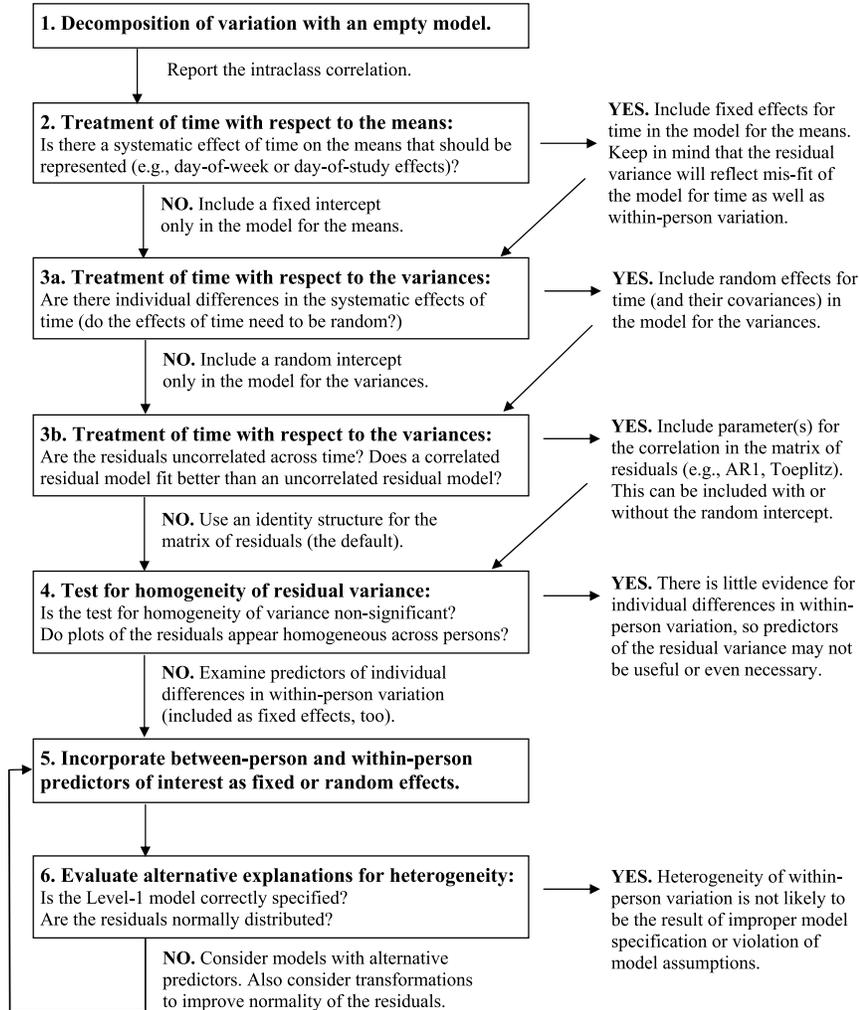
FIGURE 1   Sequence of decisions for examining heterogeneity of within-person variation.

in which y_{di} is the observed mood score for individual i on day d. The Level 1 model describes variation across days, and the Level 2 model describes variation across persons. In Level 2 model, the expected value for individual i (β_{0i}) is a function of the fixed intercept (the grand mean for the sample, γ_{00}) as well as the deviation from the fixed intercept of individual i's estimated mean over 12 days (the random intercept, U_{0i}). In Level 1, e_{di} is the remaining deviation on day d

from individual i's estimated mean. Thus, the variance of the e_{di}'s (σ_e^2) represents within-person, across day variance, known as the residual variance, and the variance of the U_{0i}'s (τ_{0i}^2) represents between-person, random intercept variance.

An intraclass correlation (ICC) can be calculated to express stability as the relative magnitude of between-person versus within-person variation. The two-level ICC is calculated as the random intercept variance divided by the total variance ($\tau_{0i}^2/[\tau_{0i}^2 + \sigma_e^2]$) and represents the proportion of variance that is between persons. The ICC for positive mood was .41 (41% variance between-persons, 59% within-persons), and the ICC for negative mood was .45 (45% variance between-persons, 55% within-persons). Thus, just over half of the overall variance was at the daily, within-person level, suggesting that individuals differed about their usual level somewhat more than they differed from each other.

Treatment of time via fixed and random effects.   It is important to note that although the data are longitudinal, a variable for time would not necessarily need to be included in Equation 1. Unlike outcomes in which systematic change and individual differences in change are expected (i.e., as in growth curve analysis; Singer, 1998), the outcomes in this example are positive and negative mood measured over 12 days. Because the data are from the baseline period of an intervention study, there is no reason why mood should change *systematically* over the 12 days, and no reason why individual differences in such change should arise. However, unintended effects of time may still be necessary to consider. For instance, positive and negative mood may differ between weekends and weekdays. In that case, a dummy variable for weekend/weekday could be included in the model as a fixed effect (i.e., differences on average between weekend days and weekdays) or as a random effect (i.e., individual differences in the difference in mood between weekend days and weekdays). This hypothesis was tested in the current data, and no significant fixed or random effects of the weekday/weekend distinction were found.

Another possibility is that the continued exposure to measurement process itself may systematically alter the outcome under study, perhaps due to the increased awareness that comes with multiple assessments. In that case, a fixed or random effect for day in study could be used to examine such reactivity. This hypothesis was also tested in the current data, and no significant fixed or random effects for day in study were found for positive mood. In contrast, however, for negative mood, there was a significant positive linear trend for day in study ($p < .05$), with marginally significant individual variation, REML χ^2 difference $(2) = 6.1$, $p < .05$, such that negative mood worsened over the course of the study. The reason for an increase in negative mood over the 12 days is unclear, although it may indicate some sort of negative continued testing effect, the extent of which varied over individuals. Although non-significant, a corresponding negative trend for day in study in positive mood (i.e., a decrease

in positive mood across days) supports this interpretation. As a result, a random linear effect of day in study will be included for negative mood to control for any possible time-dependent bias prior to examining other predictors of within-person variation.

Treatment of time via alternative covariance structures.   An equivalent way of specifying the random intercept model in Equation 1 is to specify a model with a compound symmetric structure for the variance-covariance matrix of the residuals for the 12 days (i.e., the R matrix). A compound symmetry model assumes that after accounting for differences in the estimated mean response across individuals, any remaining variation of the residuals is unsystematic across days, with no remaining residual covariation across days. Including random effects of time (i.e., between-person variances in the G matrix) results in a model that is usually more tenable for longitudinal data, in that the variances are allowed to change over time and the residuals are more correlated for time points closer together. In the absence of random effects for time, however, a compound symmetry model may still be too restrictive. In this case, alternative structures for the variances and covariances across days should also be evaluated in order to ensure appropriate tests of the fixed effects. There are many alternative structures available in SAS PROC MIXED. One such option is a first-order auto-regressive model (AR1) in which a single correlation is estimated between days that decays by a power function of r with each time lag, such that the Lag 1 correlation $=$ r, Lag $2 = r^2$, Lag $3 = r^3$, and so forth. Another option is a Toeplitz model (TOEP) in which separate correlations are estimated per lag (i.e., 11 correlations for 12 days). Each of these structures can be modified to allow variances to differ across days as needed. Finally, the variances and covariances across days can also be modeled with a combination approach in which a random intercept is estimated (in the G matrix) as well as an error correlation (in the R matrix). This combination approach posits that after accounting for variance systematic to an individual (the random intercept variance in the G matrix), residuals may still be correlated as a function of time (i.e., with a structure of AR1, TOEP, etc., in the R matrix).

The fit of alternative structures may be compared using information criteria. Because ML is known to underestimate variance components in smaller samples, REML information criteria were used instead. Akaike information criteria (AIC) and Bayesian information criteria (BIC) values in smaller-is-better forms were compared across several alternative models for each outcome. Although both values index relative model fit, the BIC also penalizes for model complexity (Singer & Willett, 2003). An unstructured matrix in which all possible variances and covariances are estimated (i.e., an unstructured R matrix) serves as the best-fitting (but least parsimonious) baseline model. For positive mood, the random intercept only (or equivalently, the compound symmetry model) had the lowest

AIC and BIC values. For negative mood, the random intercept + AR1 model had slightly lower AIC and BIC values.

Given that the random intercept only model is nested within the random intercept + AR1 model, however, their fit can be compared directly with a REML deviance test. The random intercept + AR1 model did not fit significantly better than the random intercept only model for positive mood, χ^2 difference (1) = 2.6, $p > .05$, or negative mood, χ^2 difference (1) = 3.6, $p > .05$. Allowing heterogeneous variances across days in the compound symmetry model also did not improve fit for positive mood, χ^2 difference (11) = 5.8, $p > .05$, or negative mood, χ^2 difference (11) = 17.3, $p > .05$. Thus, for positive mood, the random intercept only model provided an adequate fit to the variances and covariances. For negative mood, however, the fit of the random linear day in study model was also compared to the alternative covariance structures after including a fixed linear effect of day of study in each alternative model. REML information criteria suggested the fit of the random linear model was still preferable, and this model was used in further analyses of negative mood. The extent to which these models were still appropriate descriptors of the variances and covariances across days was re-examined after the inclusion of predictors, with similar findings.

Examining Individual Differences in the Stability of Positive and Negative Mood

The unconditional model in Equation 1 indicates that there is substantial within-person variation in positive and negative mood over the 12 days. Yet the *existence* of within-person variation over time does not necessarily imply that there are *individual differences* in the magnitude of that within-person variation. The distinction is conceptually analogous to the difference between a fixed and random effect—finding a significant effect of a predictor on average (a fixed effect) does not necessarily imply that the effect of that predictor varies over individuals (a random effect). Although the model in Equation 1 assumes that the magnitude of within-person variation over time is equivalent across persons (i.e., homogeneity of the Level-1 residual variance), this is in fact a testable assumption. Accordingly, the next issue to be addressed is whether there are individual differences in the magnitude of that within-person variation, and if so, what characteristics predict those individual differences (i.e., predictors of interindividual variation in intraindividual variation).

Testing for individual differences in the stability of mood. A test of Level-1 heterogeneity of variance (i.e., a test of the null hypothesis of no individual differences in within-person variation) is provided in the hierarchical linear modeling (HLM) program but is not currently provided in SAS PROC MIXED. However, Snijders and Bosker (1999, pp. 126–127; see also Raudenbush & Bryk,

2002, pp. 263–264) describe a method to test Level-1 homogeneity of variance that may be implemented elsewhere (see the electronic appendix for a SAS program for this test). Briefly, the method involves estimating a residual variance per person (s_i^2) using ordinary least squares regression and calculating a weighted mean of the logarithms of those residual variances, or ls_{tot} in Equation 2:

$$ls_{tot} = \frac{\sum_i [df_i \ln(s_i^2)]}{\sum_i df_i}, \tag{2}$$

where df_i = # time points—# Level 1 predictors—1,

from which a standardized residual dispersion measure, d_i, may be calculated in Equation 3:

$$d_i = \sqrt{\frac{df_i}{2}}[\ln(s_i^2) - ls_{tot}], \text{ followed by } H = \sum_i d_i^2. \tag{3}$$

The quantity H can be compared to a chi-square distribution with degrees of freedom equal to the number of individuals contributing minus 1, and a significant p-value indicates significant heterogeneity of the Level-1 residuals (i.e., significant individual differences in intraindividual variation). Raudenbush and Bryk (2002) recommend that only Level-2 units (here, persons) with $df_i \geq 10$ be included. This recommendation is based on the work of Bartlett and Kendall (1946), who suggested that the $\ln(s_i^2)$ term on which this approach is based "may safely be used for $n = 10$ and over, more tentatively from $n = 5$ to $n = 9$, and probably not at all below $n = 5$" (p. 129). Snijders and Bosker (1999) suggest an alternative approach using simulation methods if most Level-2 units have $df_i \leq 10$.

In the current example, 23 persons were included and day in study was included as a Level-1 control variable for each outcome. Not surprisingly, significant Level-1 heterogeneity of variance was found for positive mood, $H(22) = 109.39$, $p < .001$, and negative mood, $H(22) = 140.84$, $p < .001$. Yet in contrast to other contexts in which Level-1 heterogeneity may be a statistical nuisance, in this example heterogeneity manifested as differential magnitude of within-person variation is substantively interesting and is likely to represent a more realistic assumption regarding the processes under study. Indeed, why *would* one expect all individuals to exhibit the same degree of short-term fluctuation in mood?

Predicting individual differences in the stability of mood.   Given the finding of significant individual differences in within-person variation in mood,

the model in Equation 1 was extended to examine how individual differences in mental functioning (as measured by the MMSE) relate to level and stability of mood over 12 days, as shown in Equation 4:

Level 1: $y_{di} = \beta_{0i} + e_{di}$

Level 1 residual: $\sigma_i^2 = \alpha_0(\exp(\alpha_1(\text{MMSE}_i - 8) + \alpha_2(\text{MMSE}_i - 8)^2)$

Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(\text{MMSE}_i - 8) + \gamma_{02}(\text{MMSE}_i - 8)^2 + U_{0i}$

$$(4)$$

in which MMSE (centered at 8, near the midpoint of the MMSE distribution) is now included as a predictor. More specifically, in the Level-2 model for between-person variation, the expected mean level of mood for individual i across the 12 days (β_{0i}) is a function of the expected mean for a person with MMSE = 8 (the fixed intercept, γ_{00}), the additional difference in mean level of mood due to linear (γ_{01}) and quadratic (γ_{02}) effects of MMSE, and the random effect for individual i (U_{0i}). A quadratic effect of MMSE was included to test a hypothesized acceleration of the effect at higher levels of MMSE. In addition, the model for negative mood (not shown) also included an individual linear effect for day in study at level 1 (β_{1i}), which is decomposed at Level 2 into a fixed linear effect for day in study (γ_{10}) and a random linear deviation for day in study for individual i (U_{1i}).

In the Level-1 model for within-person variation, the expected value for individual i on day d (y_{di}) is a function of the individual intercept (β_{0i}) and residual deviation on day d from individual i's intercept (e_{di}). However, the variance of the Level-1 residual errors (σ_e^2) that was formerly constrained to be equal across persons is now denoted as σ_i^2 because it is now allowed to vary over individuals. Specifically, the residual variance is now separated into three pieces, as shown in the Level-1 log-linear model for the residual variance: the expected residual variance for an individual with MMSE = 8 (α_0), multiplied by the exponentiated difference in the residual variance as a function of the linear (α_1) and quadratic (α_2) effects of MMSE. The exponential function was used to normalize the variance so that a linear prediction model may be used as well as to eliminate the dependence of the variance on the mean (see Cohen et al., 2003; Littell et al., 1996; Raudenbush & Bryk, 1987, 2002).

The results from these models are given in Table 1. For positive mood, the fixed and random effects of day of study remained non-significant and were not included in the model. Although the linear fixed effect of MMSE was not significant ($p > .05$), there was a significant negative linear effect of MMSE on the residual variance, REML χ^2 difference (1) = 5.7, $p < .05$. Although overall level of positive mood was not related to mental functioning, persons with greater mental functioning reported less within-person variation in positive mood. Quadratic effects of MMSE were not significant. For negative mood,

TABLE 1
Parameter Estimates (Est.) for Multilevel Models for Stability and Covariation of
Self-Reported Positive and Negative Mood by Mental Functioning (MMSE)

Parameter	Positive Mood		Negative Mood		Covariation	
	Est.	SE	Est.	SE	Est.	SE
Intercept (γ_{00})	−0.075	0.089	−0.060	0.105	−0.185*	0.071
Day in study (γ_{10})			0.016	0.010		
Between-person negative mood (γ_{02})					−0.886*	0.154
Within-person negative mood (γ_{10})					−0.565*	0.076
MMSE fixed linear effect (γ_{01})	0.015	0.013	−0.044*	0.012	−0.034*	0.011
MMSE fixed quadratic effect (γ_{02} in Equation 4 only)			0.000	0.001		
MMSE by between-person negative mood (γ_{03})					−0.078*	0.019
MMSE by within-person negative mood (γ_{11})					−0.035*	0.013
Residual variance (α_0)	0.307*	0.025	0.300*	0.033	0.223*	0.019
Random intercept variance (τ^2_{0i})	0.215*	0.063	0.161*	0.062	0.090*	0.030
Random slope variance (τ^2_{1i})					0.048	0.030
Random intercept, random linear covariance (τ_{0i}, τ_{1i})					0.067	0.059
MMSE linear effect on residual (α_1)	−0.028*	0.011	−0.040*	0.014	−0.036*	0.011
MMSE quadratic effect on residual (α_2 in Equation 4 only)			−0.008*	0.001		
REML deviance, AIC, BIC	645, 651, 655		534, 548, 556		545, 555, 562	

*$p < .05$.

although the fixed effect of day of study was no longer significant, a significant random effect remained, so both effects were included in the model. There was a significant negative linear fixed effect of MMSE ($p < .05$) as well as significant negative linear and quadratic effects of MMSE on the residual variance, REML χ^2 difference (2) = 50.0, $p < .001$. Persons with greater mental functioning reported lower overall levels of negative mood and less within-person variation in negative mood, with an acceleration of the effect of MMSE on within-person variation at higher levels of MMSE.

The differential within-person variation across levels of mental functioning is shown in Figure 2, which plots the Level-1 unstandardized residuals against MMSE values for positive mood (top panel) and negative mood (bottom panel). As shown, after controlling for the fixed effects of MMSE (i.e., its effects on the mean), the variance of residuals is markedly smaller at higher levels of MMSE,
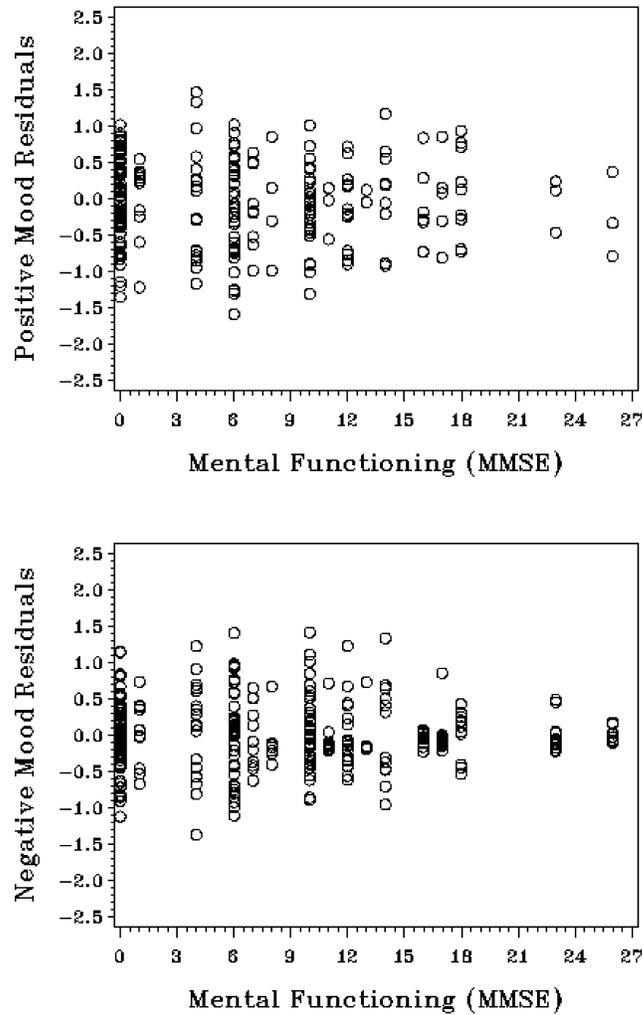
FIGURE 2 Observed unstandardized Level-1 residuals by mental functioning (MMSE) for positive mood (top) and negative mood (bottom).

particularly for negative mood. Figure 3 displays the model-predicted values for mean mood level (top panel) and magnitude of Level-1 residual variance (bottom panel) as a function of MMSE (with day held constant at 1). As shown, although greater mental functioning is related to higher mean levels of negative mood, greater mental functioning is related to greater stability in both types of mood.
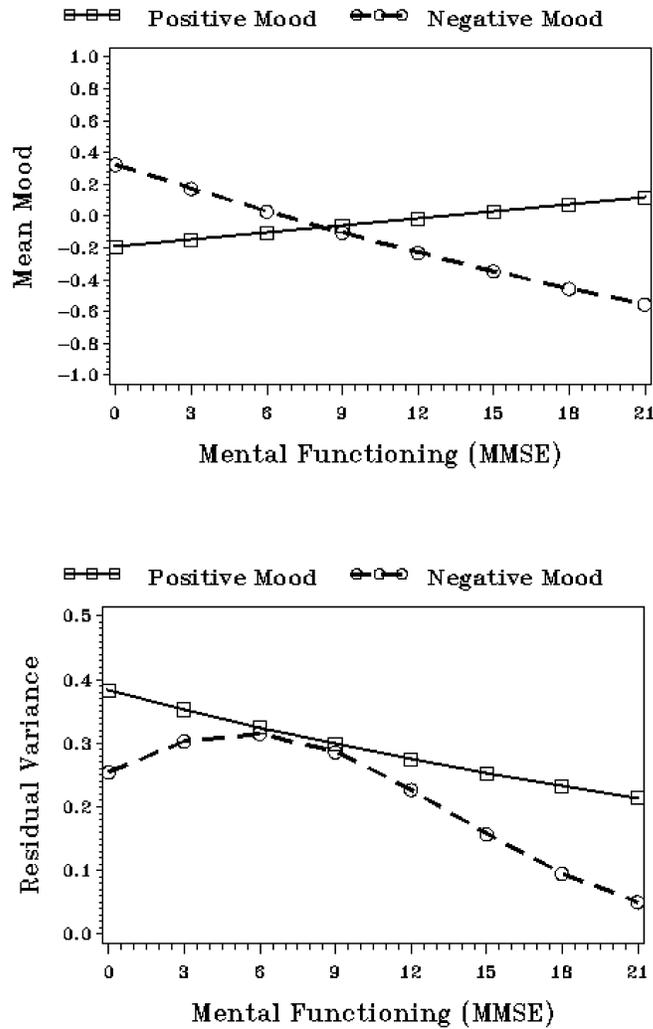
FIGURE 3 Model-predicted means levels (top) and residual variance (bottom) for self-reported positive and negative mood by mental functioning (MMSE).

Individual Differences in Covariation of Positive and Negative Mood

The models thus far have examined how MMSE relates to mean level and within-person variation over time in positive and negative mood separately. As discussed previously, however, an important issue to consider in persons with dementia is

the extent to which their self-report data exhibit internal consistency, as manifested through an inverse relationship between positive and negative mood. But because people differ in mood both from each other and from their usual levels over time, covariation of positive and negative mood should be examined at the between- and within-person levels, along with moderation of this covariation at both levels by MMSE. Greater internal consistency (i.e., more negative covariation) is expected for persons with greater mental functioning. To examine this hypothesis, the model in Equation 4 was extended to predict positive mood from negative mood as a function of MMSE, as shown in Equation 5:

Level 1: $\qquad y_{di} = \beta_{0i} + \beta_{1i}(NM_{di} - \overline{NM}_i) + e_{di}$

Level 1 residual: $\quad \sigma_i^2 = \alpha_0(\exp(\alpha_1(\text{MMSE}_i - 8))$

Level 2: $\qquad \beta_{0i} = \gamma_{00} + \gamma_{01}(\text{MMSE}_i - 8) + \gamma_{02}(\overline{NM}_i) \qquad (5)$

$$+ \gamma_{03}(\text{MMSE}_i - 8)(\overline{NM}_i) + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(\text{MMSE}_i - 8) + \gamma_{12}(\overline{NM}_i) + U_{1i}$$

in which the Level 1 model for the residual variance (σ_i^2) is interpreted as in Equation 4. The decision to have negative mood predict positive mood was admittedly somewhat arbitrary and was based on the greater amount of variation displayed in positive mood than negative mood.

In order to examine covariation at each level, the effect of negative mood on positive mood has been separated into two variables: the between-person effect, represented by the individual's observed mean across the 12 days ($\overline{NM}_i$), and the within-person effect, represented by the individual's deviation from his or her observed mean across the 12 days ($NM_{di} - \overline{NM}_i$). The between-person effect was left uncentered because zero already represented average level within the scaling of the predictor, and the within-person effect is centered relative to the individual's mean level. This separation of the between- and within-person effects of the time-varying covariate is also known as *group-mean-centering* in the multilevel modeling literature. In the Level 1 model, the response for individual i on day d (y_{di}) is a function of an individual intercept (β_{0i}) and the within-person effect of negative mood (β_{1i}). In the Level 2 model, the individual intercept (β_{0i}) is a function of a fixed intercept (γ_{00}), the main effect of MMSE (centered at 8; γ_{01}), the main effect of between-person negative mood (γ_{02}), the interaction between MMSE and between-person negative mood (γ_{03}), and an individual-specific random deviation (U_{0i}). The individual effect of within-person negative mood (β_{1i}) is a function of the fixed effect (γ_{10}), the interaction of MMSE and within-person negative mood (γ_{11}), the interaction of between- and within-person negative mood (γ_{12}), and an individual-specific random deviation (U_{1i}). Significant interactions with MMSE of between- and within-person

negative mood indicate differing magnitudes of covariation as a function of mental functioning.

The interaction of between-person and within-person negative mood was not significant, indicating that the within-person effect of negative mood on positive mood did not depend on mean level of negative mood, so this parameter was removed. The fit of the model with a random effect of within-person negative mood was marginally better than model fit with just a fixed effect of within-person negative mood, REML χ^2 difference (2) = 5.7, $p = .06$, so the random effect was retained. As before, the effects of day in study were not significant for positive mood and were not included. The results from this model are given in Table 1.

The between-person and within-person effects of negative mood on positive mood were significantly negative. Thus, as expected, mean levels and daily levels of positive and negative mood were inversely related. With regard to mental functioning, after controlling for negative mood, there was now a significant negative linear fixed effect of MMSE as well as a significant negative linear effect of MMSE on the residual variance. Thus, after controlling for negative mood, persons of greater mental functioning reported higher overall levels of positive mood but less within-person variation (greater stability) in positive mood over time. Of primary interest, however, are the significant interactions of MMSE with between-person negative mood and within-person negative mood, such that both between- and within-person covariation of negative mood with positive mood was stronger (i.e., more negative) in persons with greater MMSE. Thus, internal consistency was greater in persons with greater mental functioning, as expected. In estimating the reverse model of positive mood predicting negative mood, however, although significant negative covariation was still found both between- and within-persons, MMSE was no longer a significant moderator of those effects. Thus, this latter finding should probably be interpreted cautiously.

Heterogeneity of Variance and Model Specification

It is important to note that Level-1 heterogeneity may have many causes, including non-normality of the outcome variable or omission of a Level-1 predictor as a fixed or random effect. Raudenbush and Bryk (2002) suggest that "investigation of the possible sources of heterogeneity before concluding that a complex variance assumption is needed" (p. 263). In the current data, both outcomes were relatively normally distributed, although problems of ceiling effects (in positive mood) and floor effects (in negative mood) were observed. The Level-1 residuals were normally distributed for each, however. With regard to predictors, MMSE was included as a fixed effect at Level 2 regardless of significance in order to control for its effect on the mean prior to examining its effect on the variance. Unfortunately, to limit participant burden, very few variables were collected on a daily basis, so our possible Level-1 variables are limited to time and mood.

Nevertheless, significant heterogeneity of variance across individuals was found after considering the effect of time, and that heterogeneity was predicted by MMSE, indicating that individuals with greater mental functioning had more stable moods over 12 days (i.e., they exhibited less within-person, level-1 residual variation). The effect of MMSE on the residual variance remained significant for positive mood after controlling for overall and daily levels of negative mood. Finally, the finding of heterogeneity of Level-1 residual variance in mood is substantively interpretable as differential short-term fluctuation of mood across individuals, a more realistic scenario than the default assumption of no individual differences in fluctuation of mood. Thus, in the current example, a more complex variance function appears necessary on both empirical and theoretical grounds.

Limitations

Although the multilevel model with heterogeneous variances may be a useful analytical tool, it has limitations that should be noted. First, because heterogeneous variance models can be computationally demanding, estimation problems may be more likely. Thus the solution should be reviewed carefully in order to ensure the appropriateness of the estimates and fit statistics. It is important to note, however, that numerically different residual variance estimates between homogeneous and heterogeneous variance models are not necessarily cause for alarm. Just as main effects must be interpreted conditionally in the presence of an interaction, the intercept of the residual variance equation must be interpreted conditionally—it is the residual variance estimate *when all predictors = 0*. Thus, if 0 is not within the scale of the predictors of the residual variance, this could lead to residual variance estimates that are numerically different between models with and without predictors of variance heterogeneity but that are nevertheless estimated correctly.

An additional limitation concerns the test of heterogeneity of variance described by Snijders and Bosker (1999) and Raudenbush and Bryk (2002), which should be used cautiously for persons with $df_i \leq 10$. This restriction may be a limiting factor in studies with fewer assessments (but see Snijders & Bosker, p. 127, for an alternative procedure for assessing heterogeneity in such cases).

Finally, examination of individual differences in within-person variation is most appropriate for studies in which no systematic change is expected (i.e., studies of within-person fluctuation as opposed to within-person change). The reason for this is that the Level-1 residual variance in studies of change reflects both within-person variation and systematic mis-fit of the model of change, such that greater residual variation could result from either greater variation about the growth trajectory or from mis-fit of the growth trajectory applied to that individual. In such cases one would want to rule out systematic mis-fit before proceeding with an analysis of Level-1 heterogeneity.

SUMMARY AND CONCLUSIONS

The purpose of this work was to illustrate how multilevel models with hetero-geneous variances (i.e., dispersion models) can be used to examine individual differences in within-person variation and within-person covariation over time simultaneously. Although heterogeneous variance models are found in the ed-ucational literature, they are not commonly applied in longitudinal settings for which they could be advantageous for addressing substantive hypotheses about individual differences in the magnitude of intraindividual variability.

The use of multilevel models with heterogeneous variances for within-person variation was illustrated with self-report data of positive and negative mood over 12 days in persons with dementia. Significant within-person variation in self-reported mood was found as well as significant individual differences in within-person variation in mood (i.e., interindividual variation in intraindivid-ual variation). Greater within-person daily variation (i.e., less stability) of both positive and negative mood was observed in persons with lesser mental function-ing. Because latent traits were used as the daily outcomes, this finding is more likely to reflect true greater daily variation in mood rather than simply greater measurement error. Finally, significant negative covariation (i.e., internal consis-tency) was observed between positive and negative mood both between-persons (i.e., as covariation of overall levels) and within-persons (i.e., as covariation of daily levels), the magnitude of which may be greater in persons with greater mental functioning.

Although heterogeneity of variance can be viewed as a methodological nui-sance that one must correct for in statistical models, it can also be an interesting phenomenon in and of itself. Heterogeneity of variance may be relevant within group contexts as the extent to which individuals differ more from each other in some groups than in other groups as well as within longitudinal contexts as the extent to which some individuals fluctuate about their mean level more than do other individuals. An additional context in which these models may be useful is in studies of multiple family members. Family-level variables may be used to predict not only differences between families in level of an outcome but also differences between families in *disagreement among family members*, as indexed by Level-1 heterogeneity of variance. For instance, attitudes about gender roles may be more conservative in families with parents who are less educated, but the amount of disagreement between parents and children in attitudes about gender roles may also differ as a function of parent education. A similar approach could be used to examine differential within-group disagreement across groups in an organizational context.

The ability to examine predictors of both the mean and the variance is a unique and useful feature of the multilevel model with heterogeneous variances. Applications of this model as illustrated in this work can be useful in assisting

investigations of intraindividual variation as an important outcome in its own right. Further, the simultaneous examination of the extent to which variability is related across tasks or domains at both the between- and within-person levels is likely to provide complementary evidence to between-person approaches in evaluating common or specific determinants of behavior.

ACKNOWLEDGMENT

REFERENCES

Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods, 1*(3), 296–314.

Baltes, P., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade & P. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Bartlett, M. S., & Kendall, D. G. (1946). The statistical analysis of variances—Homogeneity and the logarithmic transformation. *Journal of the Royal Statistical Society, Supplement 8*, 128–138.

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104*(3), 396–404.

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 Associates, Inc.

Curran, P. J., Edwards, M. C., Wirth, R. J., & Hussong, A. M. (2007). The incorporation of categorical measurement models in the analysis of individual growth. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling ecological and contextual effects in longitudinal studies of human development* (pp. 89–120). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

DeShon, R. P., & Alexander, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods, 1*(3), 261–277.

Fitzmaurice, G., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. New York: Wiley.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*(4), 430–450.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting & Clinical Psychology, 68*(1), 155–165.

Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology, 14*(4), 588–598.

Kliegel, M., & Sliwinski, M. J. (2004). MMSE cross-domain variability predicts cognitive decline in centenarians. *Gerontology, 50*, 39–43.

Kolanowski, A., Hoffman, L., & Hofer, S. M. (in press). Concordance of self-report and informant assessment of emotional well-being in nursing home residents with dementia. *Journal of Gerontology: Psychological Sciences, 62B*(1), P20–27.

Kolanowski, A., Litaker, M., & Buettner, L. (2005). Efficacy of theory-based activities for behavioral symptoms of dementia. *Nursing Research, 54*(4), 219–228.

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

Lu, I. R. R., Thomas, D. R., & Zumbo, B. D. (2005). Embedding IRT in Structural Equation Models: A comparison with regression based on IRT scores. *Structural Equation Modeling, 12*(2), 263–277.

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychology and Aging, 18*(3), 510–523.

Muthén, B. O., & Muthén, L. K. (1998–2004). *Mplus user's guide* (3rd ed.). Los Angeles: Author.

Olejnik, S. F. (1998). Variance heterogeneity: An outcome to explain or a nuisance factor to control. *Journal of Experimental Education, 56*(4), 193–197.

Osgood, D. W., McMorris, B. J., & Potenza, M. T. (2002). Analyzing multiple-item measures of crime and deviance I: Item response theory scaling. *Journal of Quantitative Criminology, 18*(3), 267–296.

Overton, R. C. (2001). Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups. *Psychological Methods, 6*(3), 218–233.

Ram, N., Rabbitt, P., Stollery, B., & Nesselroade, J. R. (2005). Cognitive performance inconsistency: Intraindividual change and variability. *Psychology and Aging, 20*(4), 623–633.

Rapp, M. A., Schnaider-Beeri, M., Sano, M., Silverman, J. M., & Haroutunian, V. (2005). Cross-domain variability of cognitive performance in very old nursing home residents and community dwellers: Relationship to functional status. *Gerontology, 51*, 206–212.

Raudenbush, S. W., & Bryk, A. S. (1987). Examining correlates of diversity. *Journal of Educational & Behavioral Statistics, 12*(3), 241–269.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24*(4), 323–355.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Sliwinski, M. J., & Buschke, H. (2004). Modeling intraindividual cognitive change in aging adults: Results from the Einstein Aging Studies. *Aging, Neuropsychology, and Cognition, 11*(2–3), 196–211.

Smith, M., Gerdner, L. A., Hall, G. R., & Buckwalter, K. C. (2004). History, development, and future of the progressively lowered stress threshold: A conceptual model for dementia care. *Journal of the American Geriatric Society, 52*, 1755–1760.

Snijders, T. A. B., & Bosker, R. (1999). *Multilevel analysis.* Thousand Oaks, CA: Sage.

Strauss, E., MacDonald, S. W. S., Hunter, M. A., Moll, A., & Hultsch, D. F. (2002). Intraindividual variability in cognitive performance in three groups of older adults: Cross-domain links to physical status and self-perceived affect and beliefs. *Journal of the International Neurological Society, 8*, 893–906.

Tappen, R., & Barry, C. (1995). Assessment of affect in advanced Alzheimer's disease: The dementia mood picture test. *Journal of Gerontological Nursing, 21*, 44–46.