A Procedure for Evaluating Sensitivity to Within-Person Change: Can Mood Measures in Diary Studies Detect Change Reliably?

James A. Cranford
University of Michigan

Patrick E. Shrout
Masumi Iida
New York University

Eshkol Rafaeli
Barnard College, Columbia University

Tiffany Yip
Fordham University

Niall Bolger
Columbia University

The recent growth in diary and experience sampling research has increased research attention on how people change over time in natural settings. Often however, the measures in these studies were originally developed for studying between-person differences, and their sensitivity to within-person changes is usually unknown. Using a Generalizability Theory framework, the authors illustrate a procedure for developing reliable measures of change using a version of the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1992) shortened for diary studies. Analyzing two data sets, one composed of 35 daily reports from 68 persons experiencing a stressful examination and another composed of daily reports from 164 persons over a typical 28-day period, we demonstrate that three-item measures of anxious mood, depressed mood, anger, fatigue, and vigor have appropriate reliability to detect within-person change processes.

Keywords: *diary studies; daily mood; within-person change; reliability; Generalizability Theory*

Daily diary designs are increasingly recommended for studying dynamic psychological processes such as emotional states, self-regulation, and appraisals of social situations (Bolger, Davis, & Rafaeli, 2003; Reis & Gable, 2000). They have the potential to provide high-resolution information about evolving psychological processes, and they minimize retrospection artifacts and biases. Yet, they are not without their own methodological difficulties (Bolger et al., 2003). Diary studies impose substantial demands on participants, and the burden imposed by protocols that require frequent self-reports

Authors' Note: James A. Cranford, Addiction Research Center, Department of Psychiatry, University of Michigan; Patrick E. Shrout and Masumi Iida, Department of Psychology, New York University; Eshkol Rafaeli, Psychology Department, Barnard College, Columbia University; Tiffany Yip, Department of Psychology, Fordham University; Niall Bolger, Department of Psychology, Columbia University. Partial support for this work has come from NIH Grant T32-MH19890. We would like to thank members of the NYU Couples Laboratory for their helpful comments on drafts of this article. Correspondence concerning this article should be addressed to James A. Cranford, Addiction Research Center and Department of Psychiatry, University of Michigan, 2025 Traverwood Drive, Suite A, Ann Arbor, MI 48105-2194; e-mail: jcranfor@med.umich.edu; Patrick E. Shrout, 6 Washington Place, Mail Room 416b, Department of Psychology, New York University, New York, NY, 10003; e-mail: pat.shrout@nyu.edu; or Niall Bolger, 406 Schermerhorn Hall, 1190 Amsterdam Avenue, Department of Psychology, Columbia University, New York, NY, 10027; e-mail: nb2229@columbia.edu.

PSPB, Vol. 32 No. 7, July 2006 917-929
DOI: 10.1177/0146167206287721

may lead to biases in samples of participants toward those who are highly motivated, conscientious, and agreeable (Scollon, Kim-Prieto, & Diener, 2003). The demanding nature of diary protocols may also have negative effects on compliance (Gable, Reis, & Elliot, 2000; Litt, Cooney, & Morse, 1998).

Because of these demands, it is critical that diary measures be as brief and engaging as possible. Few standard psychological inventories can be transported wholesale into the daily diary format. As a result, researchers have relied on short forms of existing measures in an effort to reduce participant burden, and for some constructs this strategy allows for adequate coverage of the construct of interest without reducing reliability (Schimmack, 2003). Yet, development of short forms of any measure also requires attention to several important psychometric challenges (Smith, McCarthy, & Anderson, 2000). Shortened measures sacrifice redundancy that can be useful in improving reliability, and they might also restrict the conceptual range of a construct by focusing on only one or two facets of the theoretical dimension of interest.

These psychometric issues are especially critical insofar as measures included in daily diary studies are used for studying dynamic processes and change. For decades, methodologists have worried that unreliability of measures is compounded when change is studied (e.g., Cronbach & Furby, 1970; Rogosa, 1988). Systematic change variation is often small relative to random measurement error variation, and this can lead to biased inferences and loss of statistical power (Humphreys, 1996). The usual approach to improving measurement in studies of change is to use longer and more elaborate measures, but this approach is impractical in daily diary studies because of the already substantial demands on participants (Bolger et al., 2003; Reis & Gable, 2000).

In this study, we present an approach to assessing the psychometrics of short, daily diary measures based on generalizability theory (GT; Cronbach, Gleser, Nanda, & Rajaratnam, 1972). GT is an extension of classical reliability theory in that it recognizes multiple sources of variance in a given observed score. As noted by Shavelson and his colleagues (Shavelson, Webb, & Rowley, 1989), GT extends classical test theory in much the same way that factorial analysis of variance (ANOVA) extends one-way ANOVA. In one-way ANOVA, the variance in a set of scores is partitioned into between- (or systematic) and within-groups (or random) components. Likewise, classical test theory partitions the variance of an observed score into two components: true (or systematic) variance and error (or random) variance. In contrast, both factorial ANOVA and GT partition the variance of a set of scores into multiple effects and their interactions along with an error component (Shavelson et al., 1989). Although GT is a powerful psychometric framework, it is rarely applied in practice (for exceptions, see Kenny & Zautra, 2001, and Steyer, Majcen, Schwenkmezger, & Buchner, 1989, who use slightly different approaches to this question).

The GT approach (a) allows for the evaluation of traditional psychometric questions relating to reliability and validity, (b) provides a useful tool for decomposing the variance of daily measures into components that are conceptually meaningful in the context of daily process research, and (c) uses these variance components to estimate various forms of reliability, including the reliability of systematic change over time. In this article, we illustrate the usefulness of this approach for assessing the psychometric properties of short scales of daily mood.

The Study of Daily Mood

In our work on coping and support processes in intimate couples (Bolger, Zuckerman, & Kessler, 2000; Gleason, Iida, Bolger, & Shrout, 2003; Shrout, Herman, & Bolger, in press), we focus on the effects of those processes on mood, generally defined as "transient episodes of feeling or affect" (Watson & Vaidya, 2003, p. 351). In contrast to emotions, moods (a) typically last longer and are grounded in persisting temperament; (b) are readily changed by social, psychological, and environmental contexts (Clark & Watson, 1988); (c) exhibit cyclical patterns over time; and (d) refer exclusively to subjective feelings rather than other components typically associated with emotion, such as physiological responses (Watson, 2000; Watson & Vaidya, 2003). These features of mood along with their greater frequency in everyday life as compared to that of discrete emotions make them ideal candidates for the study of daily social psychological processes.

Individuals often differ in their average level of moods such as sadness, anxiety, joy, and anger, and they are also likely to experience different levels of these moods from day to day. Daily diary designs have the potential to measure fluctuations in mood in the short term as well as to construct a picture of an individual's usual range of moods. When reports of moods are averaged over many days in a wide time interval, inferences about persisting traits are possible without asking respondents to report their "usual" behavior or feelings. This aggregation strategy confers several psychometric advantages (Rushton, Brainerd, & Pressley, 1983; Watson & Vaidya, 2003).

In our work with intimate couples who are experiencing acute stress, we have limited our attention to several discrete moods that have been associated with individual functioning in the context of an acute stressor

(Bolger et al., 2000; Bolger & Zuckerman, 1995) and that also influence a variety of interpersonal processes (Thompson & Bolger, 1999). Specifically, we focus on anxious mood, depressed mood, anger, vigor, and feelings of fatigue. The inclusion of multiple moods in a diary that needed to be very brief led us to restrict ourselves to abbreviated, three-item scales for each mood. The scales were adapted from the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1992).

The Present Study

Our adaptation of the POMS both reduced the number of items to three for each scale (we excluded the Confusion-Bewilderment Scale) and changed the time frame from retrospection over the preceding week to the report of current mood. In this article we report psychometric analyses of the shortened POMS in two samples. One sample was asked to report moods in the past 24 hours, and the other sample was asked to report immediate mood at the time the diary was completed. The samples differed in another important respect. The first was recruited from a group that was experiencing a major stressful life event (Bolger et al., 2000), and the other was recruited from a population that was not experiencing a systematic stressful experience (Kennedy, Bolger, & Shrout, 2002). This feature of our design allowed us to address one aspect of the construct validity of the shortened POMS by examining trajectories of moods over time in the two samples. In addition, there is good evidence that mood varies in complex ways as a function of daily stress (Almeida, 2005; Bolger, DeLongis, Kessler, & Schilling, 1989; Suls, Martin, & David, 1998), and our inclusion of two samples provided an opportunity to assess the effects of stress on variation in within-person change.

In line with the philosophy of generalizability theory, we proposed a series of psychometric tests of the shortened POMS scales. First, we looked descriptively at the means of the five scales in each of the two samples and tracked how the means changed over the course of the diary studies. We expected to see more systematic variation in the sample of stressed persons than in the comparison sample. Based on previous work (Bolger & Eckenrode, 1991), we expected that anxious mood would be particularly sensitive to the impending stressor.

Second, we used variance decomposition methods to determine how much of the item response variation was due to systematic between-person differences in mean levels, true within-person change over time, and idiosyncratic item responses and measurement error. We expected the stressed sample to show more overall change variation because of individual differences in the response to the stressor. We used the variance component information to estimate generalizability coefficients that resemble reliability estimates for measures obtained at a single time, measures obtained by averaging ratings from multiple days, and measures of change over days.

METHOD

Overview

As mentioned, we report analyses of data from two contrasting studies, the Bar Exam Study (BES) and the Graduate Couples Study (GCS). Daily diary reports were available for 35 days in the BES sample and for 28 days in the GCS sample.

The Bar Exam Study

The BES was a study of 68 couples in which one member was preparing for the New York State Bar Examination. For more than 4 weeks leading up to and including the exam and for 3 days afterward, examinees and their partners each completed a brief daily diary questionnaire. We recruited the couples by asking officials at New York State law schools to distribute recruitment letters to their graduating students. The recruitment letter specified our inclusion criteria and stated that couples would receive $50 for their participation. From nine schools we received inquiries from 140 couples who believed they met the study's eligibility requirements. After being contacted by phone, 99 couples agreed to participate.

Approximately 2 months prior to the examination, couples were sent a background questionnaire that contained a variety of questions that are not relevant to the present study. Approximately 1 month prior to the examination, examinees and partners were each sent packets containing seven daily diary forms, which they were asked to complete each night before retiring. Participants returned their diary forms in prestamped envelopes at the end of each diary week. The diary period consisted of the 32 days leading up to and including the exam and the 3 days following the exam. We analyzed data from all 35 diary days.

Among the 99 couples who agreed to participate, 68 couples (69%) completed all of the materials. In 45 couples (66%) the examinee was male. The mean age for examinees was 29.4 years ($SD = 5.1$), and the mean age for partners was 29.5 ($SD = 5.9$). In the generalizability analyses, only the examinee reports will be used. We focused on examinees because descriptive analyses showed that the bar exam had stronger effects on their moods compared to those of their partners.

Measure of daily mood. We used an abbreviated 15-item version of the POMS (McNair et al., 1992) to assess the

five daily mood variables. Three items were selected to measure each mood. Items were selected based on (a) the magnitude of their factor loadings from a factor analysis conducted by McNair et al. (1992) and (b) the extent to which the item reflected the content domain for each mood (Smith et al., 2000). The three items measuring anxious mood were anxious, on edge, and uneasy. The three items assessing depressed mood were sad, hopeless, and discouraged. The three items tapping anger were angry, resentful, and annoyed. The three items assessing fatigue were fatigued, worn out, and exhausted. Finally, the three items measuring vigor were vigorous, cheerful, and lively. We refer to this measure as the POMS-15.

Participants were asked to indicate the extent to which they had felt or experienced these moods in the past 24 hours. They responded by circling the appropriate number on a 5-point scale ranging from *not at all* (0) to *extremely* (4). Daily scores for each mood were obtained by averaging the ratings of the relevant items. Participants had to have responded to a minimum of two items to calculate a scale mean for a given day.

The Graduate Couples Study

The GCS sample was recruited to be similar to the BES sample except that neither partner was facing a scheduled professional examination. Couples who were either married or cohabiting for at least 6 months were recruited to participate in this study. Flyers and word of mouth were used in a private urban university to inform graduate students and their friends of the study. We received 114 inquiries, and 102 couples agreed to participate. Couples were paid $50 for their participation and entered into a lottery drawing for a $1,000 prize.

The final sample consisted of 164 participants who completed at least 1 week of diaries.[1] The sample included 160 participants from couples and 4 whose partner did not complete at least 1 week of diaries. Across all participants, an average of 22.5 diary days were completed. In addition, 57% of couples were married, and the average length of cohabitation among all couples was 3.9 years ($SD = 4.1$). The average age of participants was 29.4 years ($SD = 6.4$); 52% were graduate students. As in the BES, participants were asked to complete a background questionnaire and a series of daily diary forms. The weekly diary packets consisted of seven identical, structured questionnaires to be completed at the end of each day for a total of 28 days. Participants mailed in their diaries using prestamped envelopes at the end of each diary week.

Measure of mood. Participants in the GCS sample completed an adapted POMS measure twice per day, once in the morning and again in the evening before retiring. In contrast to the BES, participants reported their current mood rather than a summary of the past 24 hours. The 15 items measuring anxious mood, depressed mood, anger, fatigue, and vigor were included within the one-page diary. We restricted our analyses of the GCS to the evening report to make the analyses as comparable to the BES as possible. Participants were asked to indicate the extent to which they were feeling or experiencing these feelings "right now, in the evening." They responded by circling the appropriate number on a 5-point scale ranging from *not at all* (0) to *extremely* (4). Evening scores for each mood were obtained by averaging the ratings of the relevant items.

Statistical Framework for GT

Generalizability analyses allow the responses by persons to a given item at a given time to be analyzed into components of variance. The variance decomposition is based on a three-way, crossed, analysis of variance model (person by day by item). For a person j, responding to mood item i on day k, the model for mood M_{ijk} is,

$$M_{ijk} = \mu + I_i + P_j + D_k + (IP)_{ij} + (ID)_{ik} + (PD)_{jk} + (IPD)_{ijk} + e_{ijk}. \quad (1)$$

The overall mean for all mood ratings of this type is μ. I_i reflects the tendency of each item i to have higher or lower scores across all persons and days, P_j reflects the equivalent effect of each person j over all items and days, and D_k is the equivalent effect of each day k across persons and items. $(IP)_{ij}$ is an effect specific to item i and person j over all days, $(ID)_{ik}$ is an effect specific to item i on day k over all persons, and $(PD)_{jk}$ is an effect specific to person j on day k over all items. The final two terms, $(IPD)_{ijk}$ and e_{ijk}, describe effects specific to item i for person j on day k. The first of these terms represents a systematic effect, and the second describes a random effect. In practice, these two terms cannot be distinguished by our design.

As a first step in the generalizability analysis, the variances associated with each of the distinguishable terms in Equation 1 are estimated from the data. Cronbach et al. (1972; also see Brennan, 2001) described this step as a "G study" analysis. In the second step, these estimates are used to make inferences about the quality of the measurements for studies with particular purposes, such as showing between-person differences or within-person changes. Generalizability theorists call the second step a "D (decision) study" analysis. Shavelson and Webb (1991) and Brennan (2001) provided detailed accounts of these steps.

When estimating the variances in the first step (the G study step), we treated all possible sources of variation

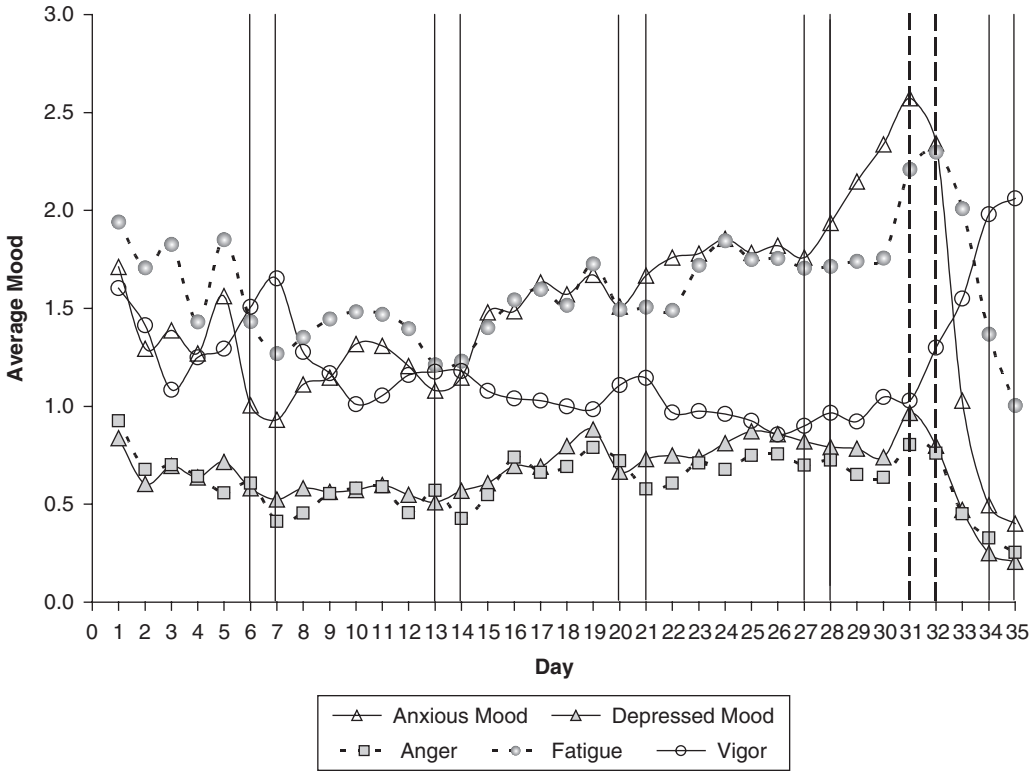

Figure 1 Trajectories of moods over 35 days among bar examinees.

NOTE: The bar examination fell on Days 31 and 32, as indicated by the dashed vertical lines. Saturdays and Sundays fell on Days 6 and 7, Days 13 and 14, Days 20 and 21, Days 27 and 28, and Days 34 and 35, as indicated by the solid vertical lines.

to be possibly random. However, in our calculations of the measurement quality of the ratings (the D study step), we only considered persons and all higher-order interactions involving persons to be random. We considered the three items for each mood and the sets of days to be fixed. These considerations allowed us to consider how reliable scales based on the specific mood measurements would be for randomly sampled persons selected at an arbitrarily chosen but fixed day.

We obtained estimates of the variance components associated with Equation 1 using the VARCOMP procedure of SAS (SAS Institute, 1997). Had there been no days with missing data, this procedure would produce exactly the same estimates as would be obtained from an analysis of the expected mean squares from the three-way, mixed model ANOVA design (Shavelson & Webb, 1991; Winer, 1971). The VARCOMP software uses as a default the MIVQUE0 method (Rao, 1971) to make use of information from the respondents who missed one or more days in the diary procedure. The MIVQUE0 method yields unbiased estimates of variance components that are invariant with respect to the model's fixed effects, and all random effects are adjusted for the fixed effects only (SAS Institute, 2004).

RESULTS

Trajectories of Moods in the BES and GCS Samples

Figure 1 shows the trajectories of anxious and depressed mood, fatigue, anger, and vigor across the 35 diary days among examinees in the BES sample. Anxious mood and fatigue stand out as the moods that were most influenced by the approaching examination. Both negative moods showed a general linear increase as the exam approached and dropped off sharply following the exam. Depressed mood and anger showed a similar pattern, albeit less pronounced. In contrast, vigor remained relatively stable and low on days leading up to the exam and then showed a sharp increase beginning on the second exam day. Weekend effects were also discernible in these data. Generally, the negative moods decreased and vigor increased on weekend

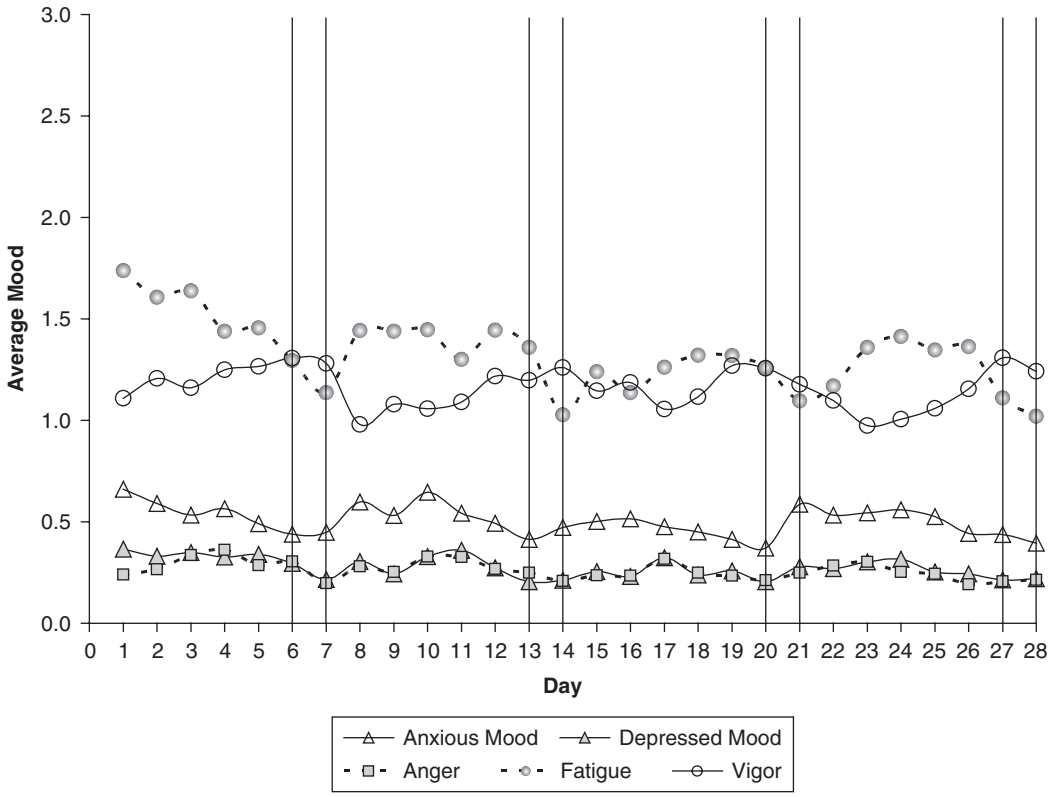

Figure 2 Trajectories of moods over 28 days among graduate students.
NOTE: Saturdays and Sundays fell on Days 6 and 7, Days 13 and 14, Days 20 and 21, and Days 27 and 28, as indicated by the solid vertical lines.

days. These weekend effects were particularly evident for anxious mood, fatigue, and vigor.

Figure 2 shows the trajectories of anxious and depressed mood, fatigue, anger, and vigor across the 28 diary days among participants in the GCS sample. Because this sample was not experiencing a systematic stressful experience, we did not expect to see the same increasing pattern of negative mood that we witnessed in Figure 1. Figure 2 shows that fatigue and vigor followed a weekly cycle, with fatigue dropping and vigor peaking during the weekend days. This weekend effect was also apparent, although not as strong, for anxious and depressed mood and anger.

Taken as a whole, these results indicated that some of the POMS-15 scales were sensitive to changes in moods that occur in the context of a major acute stressor and were also able to detect weekly cycles in moods among participants who were not faced with an upcoming stressor. We turn now to a more formal assessment of the sensitivity of the POMS-15 scales to individual differences in mood changes over time.

Variance Decomposition of the POMS-15

Within each mood scale, the item responses were assumed to vary over items, persons, and days, as described by Equation 1. The variance decomposition describes how much of the overall variance was due to each component in that equation. Table 1 shows the results of the variance decomposition analysis from the BES sample. Across all five mood scales, three components accounted for most of the variation. One is the between-person variation, which reflects whether examinees tended to have higher or lower levels of reported mood over all days and items. Another is person by day variation, which tells us that examinees had different trajectories of mood as the bar exam approached. The final large component is error variation, which was estimated by the item by person by day interaction. The impact of this last component is reduced in practice by averaging the three items in each scale.

Consistent with the trajectories shown earlier, the results indicated that anxious mood, vigor, and fatigue

TABLE 1: Variance Partitioning of POMS-15 Items in the Bar Exam Study Sample and Estimates of Between-Person Reliability and Reliability of Change

Source of Variance	*Anxious Mood*	*%*	*Depressed Mood*	*%*	*Anger*	*%*	*Fatigue*	*%*	*Vigor*	*%*
$\hat{\sigma}^2_{\text{PERSON}}$	0.383	27.1	0.251	24.3	0.158	18.1	0.521	36.8	0.255	26.9
$\hat{\sigma}^2_{\text{DAY}}$	0.218	15.4	0.018	1.7	0.013	1.5	0.066	4.7	0.077	8.1
$\hat{\sigma}^2_{\text{ITEM}}$	0.011	0.8	0.141	13.7	0.056	6.4	0.004	0.3	0.004	0.4
$\hat{\sigma}^2_{\text{PERSON*DAY}}$	0.469	33.1	0.174	16.9	0.281	32.3	0.579	40.8	0.281	29.6
$\hat{\sigma}^2_{\text{PERSON*ITEM}}$	0.041	2.9	0.117	11.3	0.073	8.5	0.019	1.3	0.057	6.1
$\hat{\sigma}^2_{\text{DAY*ITEM}}$	0.002	0.1	0.013	1.3	0.004	0.5	0.000	0.0	0.006	0.7
$\hat{\sigma}^2_{\text{ERROR}}$	0.292	20.6	0.315	30.6	0.286	32.8	0.229	16.1	0.267	28.2
Total	1.416	100	1.029	100	0.871	100	1.418	100	0.947	100

NOTE: All variance component estimates were based on 35 diary days.

TABLE 2: Variance Partitioning of POMS-15 Items in the Graduate Couples Study Sample and Estimates of Between-Person Reliability and Reliability of Change

Source of Variance	*Anxious Mood*	*%*	*Depressed Mood*	*%*	*Anger*	*%*	*Fatigue*	*%*	*Vigor*	*%*
$\hat{\sigma}^2_{\text{PERSON}}$	0.229	31.2	0.072	16.2	0.053	12.4	0.430	29.4	0.356	31.1
$\hat{\sigma}^2_{\text{DAY}}$	0.003	0.4	0.001	0.3	0.001	0.1	0.023	2.0	0.006	0.5
$\hat{\sigma}^2_{\text{ITEM}}$	0.004	0.5	0.013	2.9	0.009	2.2	0.004	0.2	0.036	3.2
$\hat{\sigma}^2_{\text{PERSON*DAY}}$	0.281	38.4	0.162	36.2	0.207	48.3	0.663	45.2	0.388	33.9
$\hat{\sigma}^2_{\text{PERSON*ITEM}}$	0.024	3.3	0.034	7.5	0.025	5.8	0.080	5.5	0.076	6.7
$\hat{\sigma}^2_{\text{DAY*ITEM}}$	0.000	0.0	0.001	0.2	0.001	0.1	0.000	0.0	0.001	0.1
$\hat{\sigma}^2_{\text{ERROR}}$	0.191	26.1	0.165	36.8	0.133	31.1	0.261	17.8	0.280	24.5
Total	0.732	100	0.448	100	0.429	100	1.461	100	1.143	100

NOTE: All variance component estimates were based on 28 diary days.

varied more with time, as indicated by the variance components for day, than the other moods in the BES. As expected, anxious mood in particular showed a large main effect of day. The person by day interaction was associated with a large proportion of variance in the items for all moods, although this effect was not as strong for depressed mood. Clearly, there were substantial individual differences in the degree to which moods changed over time in the context of a major stressor. Furthermore, these individual differences in changes over days varied across moods. For example, fatigue showed a smaller main effect of day than did anxiety, but it showed a larger person by day interaction.

The variation associated with individual items tended to be small, especially the day by item variation. One notable exception occurred for depressed mood. Here, the results showed that item and the day by item effects explained a larger proportion of the variance in the depressed mood items relative to the other POMS-15 scales. Recall that the depressed mood scale consisted of the three items sad, discouraged, and hopeless. In the context of preparing for the bar exam, perhaps the term *discouraged* was endorsed by some participants on the basis of their preparation rather than their global sense of sadness or depression. Finally, we note that the day by item variance component was close to zero for all moods, indicating that participants did not differ over days in how they responded to the POMS-15 items.

Results from the variance decomposition of the POMS-15 items for the GCS sample are presented in Table 2. With the exception of fatigue, the total variation to be explained in the mood ratings was less in the GCS sample than in the BES sample. There were however a number of similarities in the two sets of findings. In both, substantial proportions of the variance in the items for each scale were accounted for by the between-person, person by day, and error components. Consistent with theoretical arguments that mood is grounded in temperament (Watson, 2000), a relatively large proportion of variance in daily mood appeared to be trait-like.

Some noteworthy differences in the results for the BES and GCS samples were also apparent (see Tables 1 and 2). First, the main effect of day was stronger for all moods in the BES compared to the GCS sample. This effect was particularly strong for anxious mood. Second, although the person by day interaction accounted for a

TABLE 3: Generalizability Coefficients Computed from Variance Component Estimates

	R_{1F} (Between) Equation 2	R_{1R} (Between) Equation 3	R_{KF} (Between) Equation 4	R_C (Change) Equation 5
Interpretation	*Reliability (Between Persons) of Measures Taken on the Same Fixed Day*	*Reliability (Between Persons) of Measures When Persons Are Measured on Different Days*	*Reliability (Between Persons) of Average of Measures Taken Over K Fixed Days*	*Reliability of Change (Within Person)*
Bar Exam Study				
Anxious mood	.80	.41	.99	.83
Depressed mood	.73	.51	.99	.62
Anger	.66	.33	.99	.75
Fatigue	.87	.45	.99	.88
Vigor	.76	.43	.99	.76
Graduate Couples Study				
Anxious mood	.79	.41	.99	.82
Depressed mood	.60	.28	.98	.85
Anger	.58	.20	.97	.82
Fatigue	.84	.38	.99	.88
Vigor	.80	.44	.99	.81

NOTE: For the Bar Exam Study, all variance component estimates were based on 35 diary days. For the Graduate Couples Study, all variance component estimates were based on 28 diary days.

substantial proportion of variance in the items across both samples, this effect was weaker in the GCS compared to the BES for anxious mood, depressed mood, and anger. Contrary to our expectations, the person by day variance did not appear to be smaller in the GCS for fatigue and vigor. Third, the variance due to item was much higher for the depressed mood items in the BES compared to the GCS sample. This finding is consistent with the speculation that *discouraged* had specific meaning for persons studying for a difficult professional examination.

Estimation of Generalizability Coefficients

We next used the estimates of the variance components to compute generalizability coefficients that describe how reliable the POMS-15 measures would be when used in four different ways. Suppose that a researcher took a single day of reporting and computed the usual Cronbach's alpha (a measure of internal consistency) on the five scales. We define the first generalizability coefficient to be the expected between-person reliability estimate for one fixed day, symbolized as R_{1F}. It is a variance ratio composed of the between-person variation divided by the between-person variation plus estimated error variation for the scale. Using the definitions of the variances from Table 1 and defining the constant m to be the number of items (three in this case), the formula for the first coefficient is:

$$R_{1F} = \frac{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM/m}]}{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM/m}] + [\sigma^2_{ERROR/m}]}. \quad (2)$$

The numerator represents the overall expected variation in persons on the fixed set of m items. The second term of the numerator acknowledges that persons might vary in the way they respond to one or another items in the scale, but this variation is reduced by a factor of $(1/m)$ when the items are averaged to construct scales. The denominator represents the overall expected variation of the item averages (i.e., scale scores) on a single measurement day. It contains the two terms from the numerator plus a term that represents expected error variation. In neither the numerator nor denominator is it necessary to include variance terms involving time because we are assuming that all persons are measured on the same fixed day. The estimate itself uses information across all days and therefore is a kind of average of day-specific alpha coefficients across the diary days. To illustrate Equation 2, we use the results for anxious mood from Table 1. The internal consistency for a given day is calculated to be $(0.383 + 0.041/3)/(0.383 + 0.041/3 + 0.292/3) = 0.80$.

Table 3 shows the estimates of the reliability for all five mood measures when they are taken on the same day. The R_{1F} values were greater than .70 for all measures from the BES (except anger) and for three of the five measures in the GCS (except anger and depressed mood). Generally, results indicate that the ability of the three-item scales of mood states to differentiate persons on a single fixed day was moderate to good.

If different persons were to be measured on different days, then day would have to be considered to be random rather than fixed. In this case, the reliability would be lower than that given by Equation 2 because day and

person by day variation would be included in the denominator. The second estimate of reliability is based on this idea. It is represented by Equation 3 as follows:

$$R_{1R} = \frac{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM/m}]}{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM/m}] + \sigma^2_{Day} + \sigma^2_{PERSON*DAY} + [\sigma^2_{ERROR/m}]}. \quad (3)$$

The denominator of this estimate includes both σ^2_{DAY}, the average variation over days, and $\sigma^2_{\text{PERSON*DAY}}$, the variance due to the Person × Day interaction. If we were interested in estimating between-person differences in daily mood states but we had to measure different people on different days (e.g., some on weekends, some on weekdays, etc.), then this is the coefficient we would use.

Table 3 shows the estimates of the reliability, the R_{1R}, for all five mood measures in BES and GCS. The R_{1R} values were less than .50 for anxious mood, anger, fatigue, and vigor and .51 for depressed mood in BES. In the GCS, the R_{1R} for all five mood measures was less than .50. It appears that the ability of the three-item scales of mood states to differentiate persons on a single random day was poor to moderate in both samples.

The next estimate of reliability emphasizes relatively stable individual differences. Instead of treating day as random, we consider the set of days to be fixed, and we compute the average POMS score over all available days for each mood measure. Equation 4 describes the expected reliability of person-level scores computed in this way. In this equation, the number of diary days is represented by K (with $K = 35$ for the BES and $K = 28$ for the GCS). Combining over these days reduces the expected error variation by a factor of $(1/K)$.

$$R_{KF} = \frac{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM/m}]}{\sigma^2_{PERSON} + [\sigma^2_{PERSON*ITEM/m}] + [\sigma^2_{ERROR/Km}]} \quad (4)$$

Like Equation 2, Equation 4 does not include any variance terms for person by day. This version of the reliability is appropriate when considering the entire diary period as fixed.[2]

In Table 3 we see that all five scales have excellent values of R_{KF} in the BES. All values are .99. Consistent with the results from the BES sample, the between-person reliabilities (R_{KF}) in the GCS were all over .96 when the 28 days were averaged together. When it is possible to average mood ratings over a month, the means are expected to be quite stable.

The fourth and final estimate of reliability focuses on precision of the measurement of systematic change of persons over days. Here, we ask how reliable the short scales are for detecting differences in systematic changes in mood over days. Although this systematic change is considered to be noise in studies of psychological traits, it is the variance of interest in studies of process and change. This variance component, $\sigma^2_{\text{PERSON*DAY}}$, is shown in both the numerator and denominator of Equation 5.

$$R_C = \frac{[\sigma^2_{PERSON*DAY}]}{[\sigma^2_{PERSON*DAY}] + [\sigma^2_{ERROR}/m]} \quad (5)$$

As in Equations 2 and 3, the person-item-time residual term, σ^2_{ERROR}, is divided by the number of items in the scale, m, to take into account the increase in precision that results from averaging m fixed items. In Table 3, the R_C estimates suggest that systematic change in moods was reliably measured by the five scales in both samples. The lowest of these in the BES was depressed mood, which had a reliability value of .62. Furthermore, in the GCS sample, the R_C values were all above .85, indicating that the POMS-15 scales reliably measured individual differences in change over days.

DISCUSSION

The Generalizability Theory framework allowed us to address a psychometric issue of key importance for daily diary studies of psychological processes, the reliability of within-person changes. We found across two distinct samples that the POMS-15 scales showed adequate reliability for the assessment of mood changes over time. This psychometric property of the POMS-15 along with the drastically reduced number of items per scale make it a particularly useful measure for researchers interested in trajectories of change in discrete moods over time.

Although reliability of change over time was good for most POMS-15 scales, our findings indicated that the R_C coefficient for depressed mood in the BES sample was relatively low (.62). We speculate that among bar examinees, the item *discouraged* might be more salient than the other items assessing depressed mood (i.e., sad and hopeless). Consistent with this speculation, our results showed that for the BES sample, the random effect of item was indeed associated with a relatively large proportion of the variance. We had originally chosen *discouraged* over other POMS alternatives such as *blue* because it seemed to span more of the conceptual space of depressed mood, but our results suggest that it might be better to narrow the range of the items in this scale when it is being used in studies of cognitively challenging stressors.

Taken as a whole, our findings show that estimating the reliability of individual differences in change over time is psychometrically tractable and that generalizability analyses yield variance components that are statistically meaningful. Concerns over the reliability of change scores have been the subject of long-standing debate in psychology (Rogosa & Willett, 1983; Singer & Willett, 2003), and several researchers have raised important questions about the applicability of reliability estimation to the study of change (Collins, 1996). Current reliability assessment in longitudinal studies focuses on reliability estimation within waves (see Singer & Willett, 2003). Our results support the utility of applying generalizability theory to the estimation of reliability of change over time as a complement to these approaches.

Our use of a generalizability framework also allowed us to estimate the following three other forms of reliability: (a) R_{1F}, the reliability of a mood measure on an average (fixed) day; (b) R_{1R}, the reliability of a mood measure on a randomly selected day; and (c) R_{KF}, the reliability of a mood measure across all days. Reliability estimates for the POMS-15 scales for an average day (R_{1F}) were good for anxious mood, fatigue, and vigor and adequate for depressed mood and anger. As noted earlier, the values of R_{1F} can be interpreted as the average over all days of Cronbach's alpha computed for each day separately. In contrast, the reliability estimates for measures taken on a randomly selected day (R_{1R}) were low for all types of affect. This is because person by day variation is included in this coefficient as noise. One would not want to study simple individual differences in, say, vigor, and use a design that does not fix day of the week when using a measure that is sensitive to daily variation in vigor. Surveys of individual differences of vigor that are done on arbitrary or randomly determined days should use measures (other than the POMS) that are not sensitive to daily variation.

In additional exploratory analyses of the BES sample we found that the values of the day-specific reliabilities tended to be larger in the days closer to the bar examination, when there was greater variation in depressed mood and anger, than in the initial days of the diary period. In the initial days, the daily alpha coefficients were even smaller than the worrisome values of R_{1F} just reviewed. Indeed, fluctuations in scale reliability over days were reflected in the low values of the R_{1R} coefficients. Yet, when diary reports were averaged across all days, reliability estimates (R_{KF}) were excellent for all mood scales across both samples. Thus, when mood items are aggregated across days, the POMS-15 can also be used as a reliable measure of trait affectivity, but caution should be exercised when interpreting a single day's score as a trait measure.

Substantive Implications

The primary goal of this research was to evaluate the psychometric properties of the POMS-15 for the study of change, but our results also have substantive implications that are relevant for daily diary studies of mood. First, comparison of mood trajectories and variance components for the main effect of day across the BES and GCS samples showed that anxious mood, fatigue, and vigor seemed most sensitive to the upcoming bar exam. Depressed mood and anger showed little change across either sample. The results for anxious mood replicate those from an earlier diary study of medical students showing a linear increase in anxious mood as the Medical College Admissions Test exam approached (Bolger & Eckenrode, 1991).

Earlier we noted that the bar exam, as an acute, time-limited stressor, is rated as highly threatening by participants. The exam requires intense preparation, disrupts daily routines, and has profound implications for the examinee's future career prospects. Thus, it is not surprising that the exam appears to have particularly strong effects on anxious mood, vigor, and fatigue. In contrast, depressed mood and anger may be more responsive to interpersonal stressors, such as tensions and arguments with one's partner (Bolger et al., 1989; Cranford, 2004; Rafaeli, Cranford, Green, Bolger, & Shrout, 2003). A focus on interpersonal stressors may shed light on individual differences in trajectories of depressed mood and anger (Almeida, 2005).

Our results also showed weekend effects across both samples, with negative moods generally decreasing and vigor increasing on weekend days compared to weekdays. These weekend effects have also been reported in other daily diary studies of mood (e.g., Larsen & Kasimatis, 1990; Reid, Towell, & Golding, 2000; Reis, Sheldon, Gable, Roscoe, & Ryan, 2000; Stone, Hedges, Neale, & Satin, 1985). The relatively higher proportion of variance due to person by day effects in the GCS sample suggests that individual differences in the degree to which moods are entrained to a weekly cycle may be more likely to emerge in nonstressed samples. Said another way, these weekend effects may be attenuated among those coping with chronic stress. Methodologically, these weekend effects can and should be modeled in predicting trajectories of mood over several days. Theoretically, these weekend effects suggest connections between moods and basic psychological needs. For example, Reis et al. (2000) used self-determination theory (Ryan, 1995) to hypothesize that daily well-being is a function of the degree to which the basic needs of autonomy, competence, and relatedness are satisfied. In a daily process study of college students, Reis et al. found that positive affect was higher and negative affect was lower

on weekend days compared to weekdays; furthermore, daily ratings of autonomy and relatedness were higher on weekends compared to weekdays. Reis et al. suggested that weekend effects on moods may be due to greater opportunities for engaging in activities that satisfy needs for autonomy and relatedness. Our findings indicate that some forms of chronic stress may affect moods by interfering with activities that satisfy these basic needs.

Limitations, Contributions, and Directions for Future Research

There are several limitations of our study that we note. Both studies used paper-and-pencil diaries, and timing of completion cannot be objectively verified with this method (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002). Although we were not able to assess daily compliance in these studies, recent work by our group indicates that paper-and-pencil and electronic diaries yield similar results (Green, Rafaeli, Bolger, Shrout, & Reis, in press).

We only collected data on the limited set of items reported here. It is possible that other POMS items would work as well (or better) as the ones we selected on a priori grounds. We do not claim to have defined the optimal short scales based on POMS, but our empirical results suggest that the short scales we defined work well, with the caveat regarding depressed mood mentioned earlier. The POMS-15 is also limited by the shortcomings of the parent measure. The POMS only assesses one positive mood. Recent evidence strongly supports the role of positive emotions for human functioning (Fredrickson & Joiner, 2002), and other mood scales (e.g., the Positive Affect and Negative Affect Scale-X; Watson & Vaidya, 2003) could be used to assess additional positive moods.

We advocate shortening a scale such as the POMS only to make it useful in daily diary studies, where subject burden is a critical issue. We do not recommend using the POMS-15 instead of the POMS when it is possible to obtain complete information. As we noted, the trait-level information only becomes highly reliable when reports from several days are aggregated. Although we noted that aggregated data from the BES provided reliable individual difference measures, we acknowledge that these measures were obtained during a period of acute stress. Clearly the averages of anxiety and fatigue during this unique period will be higher than during normal weeks. How highly correlated these averages would be to averages taken during less stressful times is an open but interesting question.

From a psychometric point of view, our analyses are in the tradition of classical test theory (Lord & Novick, 1968). We partitioned the variance of the items over persons and days without developing an item response model that explicitly models how respondents use the five response categories. We also did not attempt to decompose the variability of responses over time according to whether it arose from autoregressive processes (e.g., as considered by Kenny & Zautra, 1995) or independent stochastic processes. Finally, we limited our analyses to describing variation within and between a priori item clusters that are simply averaged rather than considering more exploratory latent variable models of the item responses over days (e.g., Hamaker & Molenaar, 2004; Steyer, Schmitt, & Eid, 1999).

We provided results from two different studies so that we could examine the impact of diary periods that include a major stressful event. However, the measurement procedures in the two studies were not identical. In the BES, participants completed the POMS once each day and reported on their moods "in the past 24 hours." Such retrospective reports of moods, even over relatively short intervals, may be biased by more stable personality variables (Barrett, 1997). In contrast, participants in the GCS sample completed the POMS twice each day and reported on their moods "right now." Tennen and Affleck (2002) showed that momentary mood reports are not good predictors of daily moods, although we have found in our own unpublished work that momentary reports obtained at the end of the day produce results that are very similar to results obtained with instructions asking for mood "in the past 24 hours." Although we believe that the differences between the degree of daily variation in the BES and GCS studies is due to the strikingly different experience of the two samples during the diary period, we cannot rule out the alternative explanation that the different POMS instructions led to different study results.

In our analysis of the GCS, we included 160 persons whose romantic partner was also included in the study. We know that average level of moods such as reported anger are correlated as highly as .50 in our data and that daily fluctuations correlate approximately .25 (Bolger & Shrout, in press). If we were carrying out statistical tests, we would have needed to either select independent observations (e.g., by randomly selecting one couple member) or to account for the within-couples dependency to have valid tests (see Bolger & Shrout, in press). However, the goal of this analysis was to provide an example of variance decomposition of POMS reports. The fact that we had correlated participants in the GCS may have slightly reduced the size of the between person variance estimates. In fact, as N gets large relative to the size of cluster, the bias becomes negligible (Cochran, 1977). As a check, we randomly selected one member from each romantic couple and

reestimated the variance components. The results were virtually identical to those reported in Table 2.

Several strengths of our study are noteworthy. First, our results provide validation support for the classic psychometric properties of a short version of the POMS that can be used in diary studies. Also, our use of two independent samples allowed for comparison of variance components between participants who were coping with an event that should increase within-person variability in emotional states and those who were not. Thus, we could identify similarities and differences in the trajectories of moods over time that are attributable to environmental events. Finally and most important, this is the first study to our knowledge to apply a generalizability framework to assess the reliability of change. In our view, this approach has broad applicability for the development of process measures in the rapidly growing area of diary and experience-sampling research.

NOTES

1. The most frequent reason stated for withdrawing from the study was lack of time.

2. In principle, it is possible to estimate the between-person reliability coefficient for mood by averaging a randomly selected set of days (R_{KR}). Because this is not a design that has obvious advantages, the R_{KR} generalizability coefficient is not presented here, but it is available from the authors on request.

REFERENCES

Almeida, D. M. (2005). Resilience and vulnerability to daily stressors assessed via diary methods. *Current Directions in Psychological Science, 14*, 64-68.

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin, 23*, 1100-1110.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579-616.

Bolger, N., DeLongis, A., Kessler, R. C., & Schilling, E. A. (1989). Effects of daily stress on negative mood. *Journal of Personality and Social Psychology, 57*, 808-818.

Bolger, N., & Eckenrode, J. (1991). Social relationships, personality, and anxiety during a major stressful event. *Journal of Personality and Social Psychology, 61*, 440-449.

Bolger, N., & Shrout, P. E. (in press). Accounting for statistical dependency in longitudinal data on dyads. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling ecological and contextual effects in longitudinal studies of human development*. Mahwah, NJ: Lawrence Erlbaum.

Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology, 69*, 890-902.

Bolger, N., Zuckerman, A., & Kessler, R. C. (2000). Invisible support and adjustment to stress. *Journal of Personality and Social Psychology, 79*, 953-961.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.

Clark, L. A., & Watson, D. (1988). Mood and the mundane: Relations between daily life events and self-reported mood. *Journal of Personality and Social Psychology, 54*, 296-308.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley.

Collins, L. M. (1996). Is reliability obsolete? A commentary on "Are Simple Gain Scores Obsolete?" *Applied Psychological Measurement, 20*, 289-292.

Cranford, J. A. (2004). Stress-buffering or stress-exacerbation? Social support and social undermining as moderators of the relationship between perceived stress and depressive symptoms among married people. *Personal Relationships, 11*, 23-40.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74*, 68-80.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley.

Fredrickson, B. L., & Joiner, T. (2002). Positive emotions trigger upward spirals toward emotional well-being. *Psychological Science, 13*, 172-175.

Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology, 78*, 1135-1149.

Gleason, M. E. J., Iida, M., Bolger, N., & Shrout, P. E. (2003). Daily supportive equity in close relationships. *Personality and Social Psychology Bulletin, 29*, 1036-1045.

Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., & Reis, H. T. (in press). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods*.

Hamaker, E. L., & Molenaar, P. C. M. (2004, June). *The integrated state-trait model*. Paper presented at the annual meeting of the Psychometric Society, Pacific Grove, CA.

Humphreys, L. G. (1996). Linear dependence of gain scores on their components imposes constraints on their use and interpretation: Comment on "Are Simple Gain Scores Obsolete?" *Applied Psychological Measurement, 20*, 293-294.

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology, 63*, 52-59.

Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 243-263). Washington, DC: American Psychological Association.

Kennedy, J. K., Bolger, N., & Shrout, P. E. (2002). Witnessing interparental psychological aggression in childhood: Implications for daily conflict in adult intimate relationships. *Journal of Personality, 70*, 1051-1077.

Larsen, R. J., & Kasimatis, M. (1990). Individual differences in entrainment of mood to the weekly calendar. *Journal of Personality and Social Psychology, 58*, 164-171.

Litt, M. D., Cooney, N. D., & Morse, P. (1998). Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology, 17*, 48-52.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *EdITS manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

Rafaeli, E., Cranford, J. A., Green, A. S., Bolger, N., & Shrout, P. E. (2003, February). *The good and bad of relationships: Effects of social hindrance and social support on relationship moods in daily life*. Poster session presented at the annual meeting of the Society for Personality and Social Psychology, Los Angeles.

Rao, C. R. (1971). Estimation of variance and covariance components—MINQUE theory. *Journal of Multivariate Analysis, 1*, 257-275.

Reid, S., Towell, A. D., & Golding, J. F. (2000). Seasonality, social zeitgebers and mood variability in entrainment of mood: Implications for seasonal affective disorder. *Journal of Affective Disorders, 59*, 47-54.

Reis, H. T., & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reiss & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190-222). Cambridge, UK: Cambridge University Press.

Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin, 26*, 419-435.

Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-209). New York: Springer.

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, *20*, 335-343.

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, *94*, 18-38.

Ryan, R. M. (1995). Psychological needs and the facilitation of integrative processes. *Journal of Personality*, *63*, 397-427.

SAS Institute. (1997). *SAS/STAT software: Changes and enhancements through release 6.12*. Cary, NC: Author.

SAS Institute. (2004). *SAS/STAT 9.1 user's guide*. Cary, NC: Author.

Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies*, *4*, 79-106.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies*, *4*, 5-34.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, *44*, 922-932.

Shrout, P. E., Herman, C., & Bolger, N. (in press). The costs and benefits of practical and emotional support on adjustment: A daily diary study of couples experiencing acute stress. *Personal Relationships*.

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

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 102-111.

Steyer, R., Majcen, A. M., Schwenkmezger, P., & Buchner, A. (1989). A latent state-trait anxiety model and its application to determine consistency and specificity coefficients. *Anxiety Research*, *1*, 281-299.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*, 398-408.

Stone, A. A., Hedges, S. M., Neale, J. M., & Satin, M. S. (1985). Prospective and cross-sectional mood reports offer no evidence of a blue Monday phenomenon. *Journal of Personality and Social Psychology*, *49*, 129-134.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *British Medical Journal*, *324*, 1193-1194.

Suls, J., Martin, R., & David, J. P. (1998). Person-environment fit and its limits: Agreeableness, neuroticism, and emotional reactivity to interpersonal conflict. *Personality and Social Psychology Bulletin*, *24*, 88-98.

Tennen, H., & Affleck, G. (2002). The challenge of capturing daily processes at the interface of social and clinical psychology. *Journal of Social and Clinical Psychology*, *21*, 610-627.

Thompson, A., & Bolger, N. (1999). Emotional transmission in couples under stress. *Journal of Marriage and the Family*, *61*, 38-48.

Watson, D. (2000). *Mood and temperament*. New York: Guilford.

Watson, D., & Vaidya, J. (2003). Mood measurement: Current status and future directions. In J. A. Schinka, W. F. Velicar, & I. B. Weiner (Eds.), *Handbook of psychology, Vol. 2: Research methods in psychology* (pp. 351-375). New York: John Wiley.

Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

Received August 4, 2005
Revision accepted January 12, 2006