Variability in intensively assessed mood: Systematic sources and factor structure in outpatients with opioid use disorder

Samuel W. Stull[1,2], **Jacqueline Mogle**[3], **Jeremiah W. Bertz**[2], **Albert J. Burgess-Hull**[2], **Leigh V. Panlilio**[2], **Stephanie T. Lanza**[1,3], **Kenzie L. Preston**[2], **David H. Epstein**[2]

[1]Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, 16802, USA

[2]Intramural Research Program, National Institute on Drug Abuse, 251 Bayview Blvd., Suite 200, Baltimore, MD, 21224, United States

[3]The Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University, University Park, PA, 16802, USA

Abstract

In intensive longitudinal studies using ecological momentary assessment (EMA), mood is typically assessed by repeatedly obtaining ratings for a large set of adjectives. Summarizing and analyzing these mood data can be problematic because the reliability and factor structure of such measures have rarely been evaluated in this context, which—unlike cross-sectional studies—captures between- *and* within-person processes. Our study examined how mood ratings (obtained thrice daily for 8 weeks; n = 306, person moments = 39,321) systematically vary and covary in outpatients receiving medication for opioid use disorder (MOUD). We used generalizability theory to quantify several aspects of reliability, and multilevel confirmatory factor analysis (MCFA) to detect factor structures within and across people. Generalizability analyses showed that the largest proportion of systematic variance across mood items was at the person level, followed by the person-by-day interaction and the (comparatively small) person-by-moment interaction for items reflecting low arousal. The best-fitting MCFA model had a three-factor structure both at the between- and within-person levels: positive mood, negative mood, and low-arousal states (with low arousal considered as either a separate factor or a subfactor of negative mood). We conclude that (1) mood varied more between days than between moments, and (2) low arousal may be worth

scoring and reporting separately from positive and negative mood states, at least in a MOUD population. Our three-factor structure differs from prior analyses of mood; more work is needed to understand the extent to which it generalizes to other populations.

Keywords

multilevel confirmatory factor analysis; generalizability theory; ecological momentary assessment; mood; affect; opioid use disorder

Introduction

Investigators now routinely use experience-sampling methods such as ecological momentary assessment (EMA) to capture experiences as they unfold in near real time (e.g., Shiffman, 2009). This approach enables the temporal examination of a person's self-reported experience with a particular event (e.g., mood before a drug-use event) to be examined when and where it occurs. The data derived from EMA, sometimes called intensive longitudinal data (ILD), can then be examined with statistical methodology such as multilevel modeling to help disentangle within-person associations from associations at the between-person level (Hox et al., 2017). This exciting new approach has aided in uncovering within-person associations of self-reported experiences and health outcomes, and has simultaneously raised fundamental questions about how to optimally develop EMA measures, which must be brief yet still capture moment-to-moment variability reliably (McNeish et al., 2021). Two analytical techniques, multilevel confirmatory factor analysis (MCFA) and generalizability theory (G-theory), are ideal for informing measurement development in ILD, because they elucidate the factor structure and quantify the reliability of instruments from which ILD data are collected. The present investigation highlights the utility and complementary aspects of these two analytical techniques and simultaneously provides new substantive insights into the assessment of mood among people with opioid use disorder (OUD). We begin with background on the assessment of mood, including among people with addictions, and on G-theory and MCFA.

Assessment of mood

Moods[1] are central to human experience and are associated with cognitions, health, and health behaviors (e.g., Burgess-Hull & Epstein, 2021; Denollet & De Vries, 2006; Forgas, 2008; Steptoe et al., 2005)—yet much about mood remains controversial, including how to conceptualize its components. One of the most widely used conceptualization—which is sometimes described as the mood circumplex—posits that mood experiences exist along two continuums: a low-arousal to high-arousal continuum, and an unpleasant to pleasant continuum (valence) (Russell, 1980). Moods are also conceptualized and operationalized based on positive and negative valence only (Watson et al., 1988). Regardless of the

---

[1]In this paper, we use the terms *mood* and *affect* interchangeably. Both terms refer to valenced emotional states. Some investigators distinguish mood from affect; the nature of the distinction can vary, but often it reflects a conceptualization of affect as an outwardly observable state and mood as a subjectively reported state (Serby, 2003). All data discussed in this paper (including those from measures such as the Positive and Negative Affect Schedule) refer to subjectively reported states. Nonetheless, we use the abbreviations *PA* (Positive Affect) and *NA* (Negative Affect) because they have become standard in literature on mood.

specific assessment tool, or specific item(s), mood is typically assessed as a state or as a trait tendency toward particular mood states. These tendencies in some cases can inform psychiatric diagnoses (e.g., major depressive disorder, bipolar I and II disorders) and provide information regarding patient well-being and risk for worsening symptomology. For addiction, for example, mood states are not explicitly part of the of the diagnostic criteria, but they can serve as important indicators of patient experience and inform risk or protective factors in addiction recovery (Panlilio et al., 2019; Preston et al., 2018a; Preston et al., 2018b).

Instruments used to capture mood in the context of ILD, however, have been informed almost exclusively by questionnaires developed to assess mood on a single measurement occasions; this measurement occasion is typically in a laboratory or clinical outpatient settings (e.g., Shacham, 1983; Watson et al., 1988). The measurement properties of such mood-assessment tools have received very little scrutiny in the context of ILD, but we believe that such work is greatly needed to disambiguate sources of error from systematic variance that is substantively informative at the within- and between-person levels. Specifically, understanding how variance in mood may be captured with EMA study-design units of measurement (persons, days, and moments). In addition, capturing the factor structure of mood at the within- and between-person levels from EMA data can help inform a parsimonious selection of items that best capture mood valence, and will determine whether specific aspects of mood are better assessed as states, traits, or both. Together, understanding mood variance across units of measurement and its factor structure may help to optimize the ambulatory assessment of mood in many populations, including people with OUD.

Generalizability theory

Generalizability theory (G-theory; (Shavelson & Webb, 1991)) is a statistical approach used for differentiating variance in measurement scores based on underlying facets (i.e. sources of underlying variability). Facets can be operationalized as features of the study design such as people, items, or measurement occasions. Importantly, G-theory also specifies interactions among these facets, reflecting effects of one facet that depend on the level of another facet. For example, a person's response to a particular item may depend on person-specific characteristics (e.g., age, sex) and when the item is presented (e.g., weekend vs. weekday). G-theory offers the major advantage of differentiating each facet's contribution to measurement error, whereas, in classical test theory, measurement error is a sum of all sources of error irrespective of its source.

G-theory has been helpful for developing measures that differentiate between states and traits, as has been shown, for instance, in the measurement of anxiety and mindfulness (Forrest et al., 2021; Medvedev et al., 2017). G-theory has also been used with daily-diary data to help understand how variance in mood scores is attributable to differences in people, days, items, and their interactions (Scott et al., 2020). Daily-diary designs offer the unique benefit of capturing changes in mood states across days and capturing stable between-person patterns perhaps analogous to mood traits. Between-person mood scores can be derived from daily-diary data based on averages of mood from many assessment occasions; this approach

helps minimize recall bias and has the psychometric benefit of more reliable between-person estimates (Cranford et al., 2006). Among people with addiction, clarifying mood traits can inform *who* is at elevated risk for problematic outcomes such as drug craving or substance use, and can identify the moments or days *when* the outcomes are more likely. However, a critical, but often overlooked step before investigating covariance between moods and outcomes is to first understand how variance in mood items manifests across different units of measurement commonly assessed in EMA studies: people, days, moments, and their interactions. For example, if variance is greater between days than between moments, this may suggest fluctuations in mood are more closely linked with the day of the week or day of the study than a particular moment in a day. Additionally, if some items show little variability across moments or days, that might suggest an insensitivity that would argue against using the item in future studies.

Multilevel factor analysis

As discussed previously, the development of mood measures has been almost exclusively based on cross-sectional data. These measures provide a single assessment of a mood state or a general assessment of mood-related traits; they do not allow for the statistical disentangling of state from trait and can conflate the two. Although useful for between-person investigations, cross-sectional designs do not permit estimates of within-person variability; further, between-person estimates of traits are confounded with within-person variability (Hoffman & Stawski, 2009). To appropriately capture and separate variance from both mood states and traits, measures need to be appropriate for intensively repeated administration. In ILD, the between-person structure indicates how item clustering differentiates individuals from one another, whereas the within-person structure indicates how item clustering may differentiate moments (or days) from one another. Only a handful of studies have explored the factor structure of mood in repeated-measures data.

Exploration of mood factor structure with repeated measures has already begun for one frequently used instrument, the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988), with two investigations using multilevel confirmatory factor analysis on daily-diary PANAS data from nonclinical adult samples. One daily-diary study, with five PANAS measurement occasions, found correlated positive and negative mood factors at the between- and within-person day levels (Merz & Roesch, 2011). Another study, which included two samples administered the PANAS on 7 or 14 occasions, also found positive and negative mood factors at the between- and within-person day levels (Rush & Hofer, 2014). These investigators found that positive and negative mood were inversely correlated at the within-person day-level but independent at the between-person level. This was an important step forward for measuring mood with ILD, but the factor structure of mood should be investigated across more measurement occasions and in other samples, particularly where there is reason to believe that mood variance and/or factor structure may be different for people with psychiatric disorders (e.g., Trull et al., 2008). For example, people with addictions may have elevated negative mood and reduced positive mood, with important implications for the variance and covariance of mood items (Baker et al., 2004; Garfield et al., 2014; Stull et al., 2021).

Particular aspects of mood may be more salient for people with addiction, such as the intensity or frequency of low- or high-energy moods. Apart from the psychomotor stimulant or sedative effects of drug use itself, people with psychiatric disorders sometimes deny experiences of negative moods but report feelings of fatigue or malaise; sometimes lower positive moods can be mistaken or conflated with feelings of lethargy (Billones et al., 2020; Chen, 1986). For these reasons it is important to more fully understand the factor structure of mood for people with addictions, including those with OUD, because issues with reduced positive moods and/or lower energy in this population may be especially prominent (Garfield et al., 2014; Stull et al., 2021) and may confer risk for negative outcomes such as return to use or addiction treatment dropout (Panlilio et al., 2019).

Study Aims

The present study has two aims: (1) to apply G-theory to capture systematic between-person and within-person sources of variance in mood at three nested levels (person, day, and moment of assessment within day) among people with opioid use disorder and (2) to apply multilevel confirmatory factor analysis (Muthén, 1991; Preacher et al., 2010) for examination of the factor structure of mood in this population at each relevant nested level.

Method

Sample

The sample consisted of 306 participants who were receiving outpatient treatment with opioid agonist medication (51% receiving methadone, 49% receiving buprenorphine-naloxone). The participants were predominately male (78%) and African American (64%) (about a third were white (34%) and a small proportion were Asian or more than one race (2%). A small proportion of the sample (2%) were Hispanic or Latinx. The average age was 48 years (SD = 10.0). Greater detail regarding the participant clinical and demographic characteristics can be found elsewhere (Panlilio et al., 2021). Eligibility criteria were: age 18–75 years; physical dependence on opioids; and residence in or near Baltimore, Maryland. Exclusion criteria were: any DSM-IV psychotic disorder; history of bipolar disorder; current major depressive disorder; current DSM-IV dependence on alcohol or sedative-hypnotics; cognitive impairment severe enough to preclude informed consent or valid self-report; or medical illness that would prevent participation. All study procedures were approved by the Addictions Institutional Review Board of the National Institutes of Health.

Procedure

Data were drawn from an observational ecological momentary assessment (EMA) study investigating the relationships among mood/affect, stress, craving, and drug use in patients in outpatient treatment for opioid use disorder (OUD). (The study was registered as a clinical trial in accordance with NIH requirements; National Clinical Trial Identifier NTC00787423.) The study was a long-term project, with enrollment from July 2009 through April 2018, though each participant was only enrolled for up to nine months. Recruitment took place in three waves, where participants received opioid agonist medication (naloxone-buprenorphine or methadone) from the study treatment-research center or at a local community treatment clinic. Each wave differed in their length of time in the study, but

the data presented here are from the first eight weeks (56 days) of outpatient treatment. All participants who completed at least two EMA responses were included in analysis. Participants visited the treatment-research clinic two or three times per week to provide urine samples for drug testing, receive counseling for OUD, and for compliance checks with the smartphones that were provided. Participants completed questionnaires delivered to their smartphones three times per day at randomly scheduled times during their normal waking hours (i.e., "random prompts"). Research staff met with participants weekly to answer questions about smartphone-based assessments and to review compliance for the prior week. If participants had missed no more than three random prompts that week, they received $10-$30 (depending on the study wave). The mean number of random prompts completed during the first 8 weeks of outpatient treatment was 126.90 (SD=43.42). Nearly 80% of participants completed EMA reports throughout the first 8 weeks of outpatient treatment and only 4% of participants provided less than a week of EMA reports. Overall, the mean proportion of random prompts completed was .82, (SD =.14), with participants completing, on average, 2.46 prompts per day (SD=.42). Participants also self-initiated reports following drug use and more-than-usual stress and completed end-of-day reports. None of these assessments were analyzed in the present study; they have been the subject of prior publications (Burgess-Hull et al., 2021; Furnari et al., 2015; Preston et al., 2017; Preston et al., 2018b).

Measures

Demographic and Clinical Assessment—Upon joining the study, participants completed version 5 of the Addiction Severity Index (ASI) (McLellan et al., 1985), a semistructured interview that assessed demographic information (e.g., sex, race, ethnicity, and age) and history of drug use, mental-health issues, criminal-justice involvement, social support, and employment.

Ecological Momentary Assessment—In the smartphone-based random prompts, participants were asked to rate their mood/affect for how they felt "just before the electronic diary beeped" based on 24 affect adjectives adapted from the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988). These items were intended to assess a broad range of positive and negative affect. Positive affect (PA) items were *feel like celebrating, happy, lively, cheerful, relaxed, vigorous, contented, pleased,* and *carefree;* negative affect (NA) items were *anxious, fatigued, worn out, annoyed, sleepy, afraid, discouraged, resentful, hopeless, bored, uneasy, sad, exhausted, angry,* and *on edge.* Responses ranged from 1 (*not at all*) to 5 (*extremely*).

Statistical Analyses—We computed means and percentages to describe the clinical and demographic characteristics of the sample and estimated the person-level means and overall means and standard deviations of each PANAS item. Next, we computed the intraclass correlation coefficients (ICCs) for each PANAS item using separate linear mixed models. The ICC captures the proportion of between-person variability relative to the total variability. To address our first aim of identifying systematic sources of variance due to study design variables, we used G-theory (Shavelson & Webb, 1991). G-theory is analogous to a factorial analysis of variance (ANOVA) in that variation is partitioned to

underlying sources, also referred to as facets, which refer to features of a study design. The current study included facets for person, day number (day categories of 0-56 for each person), and moment (moment categories 1-3 from random prompts within each day). In our results, we interpret these as the amount (and relative proportion) of variance explained by each facet. For example, the main effect for the momentary facet captures whether ratings of mood consistently depended on it being the first, second, or third random prompt of the day (across all persons and days). A Venn diagram of the mood-variance decomposition consistent with our analyses is provided in Figure 1. The center of the Venn diagram represents the three-way interaction of person, day, and moment, along with any remaining variance unexplained by the main effects or interactions of the design features. In preliminary analyses, we found that the variance explained by the three-way interactions for all individual mood items was negligible. Because these three-way interactions were small and confounded with the error term, we simplified our models by removing this term. Thus, items were modeled separately with the following predictors: person, moment, day, and all possible two-way interactions. Variance in the mood items not explained by our predictors included random error variance, and likely systematic sources of variance other than the facets we included in our models. Analyses were conducted using PROC VARCOMP with the MIVQUE0 method, which allows variance-decomposition models to be estimated with missing data on outcomes assumed to be missing at random (SAS Institute, 2014).

To address our second aim of examining the factor structure of the affect items both between persons and within persons across moments, we used multilevel confirmatory factor analysis (MCFA), which allows the factor structure at multiple levels to be estimated simultaneously (Muthén, 1991; Preacher et al., 2010). At each level, factors were allowed to correlate unless specified otherwise.

$$Y_{ijk} = \nu + \lambda_{wm}\eta_{ijk} + \varepsilon_{ijk} + \lambda_{wd}\eta_{ij} + \varepsilon_{ij} + \lambda_b\eta_i + \varepsilon_i$$

where Y_{ijk} is a p-dimensional vector of the assessed item for a given individual i on a given day j and moment k; p denotes the number of indicators used (i.e. affect items) and v is the p-dimensional vector of intercepts. Let q represent the number of latent variables. λ_{wm} represents the p x q momentary-level factor loadings matrix, λ_{wd} represents the p x q within-person day-level factor loadings matrix, and λ_b represents the p x q between-person factor loading matrix. Further, η_{ijk} represents the momentary-level latent variable, η_{ij} represents the day-level latent variable, and η_i represents the person-level latent variable. Variance for the p-dimensional vectors unique to each indicator is represented as ε_{ijk} for the momentary-level factor, ε_{ij} for the day-level factor, and ε_i for the person-level factor.

To quantify the discrepancy between the estimated model and the observed data, we used several model-fit statistics: the chi-square test statistic (χ^2), the Tucker-Lewis index (TLI), the root-mean-square error of approximation (RMSEA), and the standard root-mean-square residual (SRMR). We did not rely heavily on the chi-square statistic for model evaluation because it can be particularly sensitive to small deviations in model fit with large samples (Kline, 2016). We used the TLI and RMSEA as indicators of model fit overall, and the SRMR for model fit separately at the different levels. For the TLI, values >0.90 indicate

reasonable fit given large samples. For RMSEA and SRMR, values <0.08 are commonly used as indicators of adequate model fit. For more on the performance of fit indices given the sample size and the consideration of multiple fit indices simultaneously, see Hu and Bentler (1999). We used these fit indices along with theoretical reasoning outlined in the introduction (e.g., circumplex vs. valence only) to evaluate CFA models with 2-4 factors at the within- and between-person levels.

The affect indicators were Likert-style questions with five possible response options, which we treated as continuous in all analyses. We attempted to fit multilevel factor models with Likert responses treated as ordinal, but had to forgo that approach due to software and computation limitations. Such limitations are not uncommon when fitting complex multilevel factor models with ordinal data, which can lead to unstable factor solutions (Grilli & Rampichini, 2007). Prior simulation studies of confirmatory factor analyses have demonstrated that ordinal items with five or more response options generally yield comparable results whether treated as ordinal or continuous (Rhemtulla et al., 2012). Ultimately, our decision to treat Likert items as continuous was consistent with the approach taken in two prior MCFAs of the PANAS (Merz & Roesch, 2011; Rush & Hofer, 2014); this came with the added benefit of making our results more easily comparable to this prior work.

This study was part of a larger observational study (CTN NCT00787423; internal protocol number 09-DA-N020), but the specific design and hypotheses of this particular study were not preregistered. The data and study materials are available upon request from the author.

Results

Descriptive Information and ICCs

Individual PA items generally had larger means and standard deviations than NA items, ranging from the highest mean for relaxed (M = 2.11, SD = 0.76) to the lowest mean for afraid (M = 1.09, SD = 0.20) (see Figure 2; Tables 1 and 2). ICCs were greater for all PA items compared to NA items, ranging from the highest for pleased (0.58, indicating that most of the variability was between-person) to the lowest for angry (0.21, indicating that most of the variability was within-person) (Tables 1 and 2 Correlations of mood items were investigated as well, with clusters of mood items tending to occur for NA and PA at both person- and momentary-levels. Low-arousal mood items (e.g., "sleepy", "fatigued") were also highly correlated, particularly at the momentary level. (See Supplemental Figures 1 and 2 for person- and momentary-level correlation matrices).

Variance Decomposition of EMA PANAS Items

Results for the PANAS variance decomposition, shown as the proportion of variance explained by each facet (persons, days, moments and their two-way interactions) are presented in Table 1 for PA items and Table 2 for NA items. Most apparent is the large proportion of person-level variance among all PANAS items: differences across participants explained between 49.2% and 60.9% of the variance of the PA items, and between 17.4% and 39.4% of variance for the NA items. The percent of variance explained by the person

by day interaction was the second largest for all mood items (Table 1 and Table 2). This suggests that, within person, fluctuations across days explain between 9.8% (*cheerful*) and 22.8% (*sad*) of the variance in mood (above and beyond person alone, and above and beyond any person-invariant trend across days in the study). The remaining facets (day, moment, day*moment, and person*moment) typically explained <1% of variance in the mood items. A notable exception was the person by moment interaction for the items that captured low-arousal moods, including *sleepy, worn out, tired,* and *exhausted.* Within persons, the variance in low-arousal items varied by moment of the day. The highest percentage of unexplained variance was for the item *angry* (64%); the lowest was for *carefree* (28%). (See Online Supplemental Figure 3 for a graph of the variance decomposition for each mood item.)

Multilevel confirmatory factor analysis (MCFA) of EMA PANAS items

We hypothesized two separate, correlated latent factors—PA and NA—at the person-, day-, and momentary- levels (Merz & Roesch, 2011; Rush & Hofer, 2014). Because our data had three levels of nesting (moments nested within days, and days nested within participants), we first fit an MCFA model with a three-level factor structure (Model 1). We specified two factors, one for PA and one for NA with simple structure (i.e., no cross-loadings of items onto both factors) at each level. Model-fit indices were overall quite poor, with a particularly poor SRMR fit statistic at the day level (0.384).

We next fit an MCFA model in which we still accounted for the three-level structure, but aimed to improve the fit by reducing complexity. This model again included two correlated factors (PA and NA), each specified at the person level (level 3) and the moment level (level 1) but not the day level (level 2), though we continued to account for correlations at the day level. The model-fit indices again were generally poor (Table 3; Model 2) and the SRMR between-level statistic for day-level was large (0.121), which again suggested that accounting for the nesting structure at the day-level may not be advantageous for improving model fit.

For our third model, we removed the day-level structure and simply modeled moments nested within people (Table 3; Model 3). Because model fit did not substantially worsen, we chose to remove the day-level factor model for subsequent models in the interest of parsimony. Loadings onto the NA factor were low for *tired, sleepy, exhausted,* and *fatigued,* particularly at the within-person level, suggesting that a separate factor might be needed to explain the variance in these low-arousal items, consistent with theory described in the introduction.

Model 4 tested a correlated, simple three-factor structure with PA, NA, and low-arousal factors at the within-momentary and between-person levels (with no cross-loadings specified), substantially improving all fit indices (Table 3; Model 4).[2]

---

[2]Because prior studies have shown factor analysis is sometimes inadequate for capturing bipolar factors, we performed diagnostic checks outlined by van Schuur and Kiers, 1994 to test for this possibility. Our diagnostic checks did not provide evidence that our factor analysis obscured bipolar factors.

We then further refined Model 4 (Table 3; Model 4a), first to align with prior theoretical work and account for unique covariances between particular items. Specifically, items from the low-arousal factor were specified to cross-load on the NA factors (and these factors were uncorrelated) at both the between-level and person-level factor structure (Table 3; Model 4a) to be consistent with prior theoretical work suggesting that stress and fatigue may be separate but interrelated factors (Denollet & De Vries, 2006). Thus, the low-arousal factor was specified as a subfactor of NA (with the NA and Low Energy factors specified to be uncorrelated). Specifying low energy as a subfactor was important conceptually for allowing its items to load onto the NA factor while also allowing it to remain distinct to account for the unique covariance among low-arousal items. This updated model resulted in a negligible improvement in model fit, but had the advantage of greater interpretability.

We next inspected the modification indices of Model 4a to see whether any particular items shared unique variance not specifically explained by the factors. We found that *angry* with *annoyed* for the NA factor, and *pleased* with *contended* for the PA factor, had substantially large modification indices at the within-person level (MI = 983.403 and 632.249, respectively); thus, these residual covariances were added in Model 4b. Fit of Model 4b improved further (TLI = 0.902, RMSEA = 0.017, SRMR between = 0.059, and SRMR within = 0.037). Thus, we selected Model 4b to represent the final multilevel factor structure. The factor structure and standardized factor loadings for Model 4b are shown in Figure 3. Finally, for completeness, we estimated a 4-factor model to test whether the low-arousal factor might reflect activation/deactivation consistent with the four quadrants of a circumplex model of affect, with four unipolar factors at the within-momentary and between-person levels. The four factors were low- and high-arousal PA and low- and high-arousal NA. The fit statistics for this model were the poorest of all models tested (Table 3; Model 5).

Discussion

For our G-theory investigation, we studied the individual items to explore how facets of our design might explain unique variance within each mood item. Our initial descriptive analyses indicated that some items varied differently from one another even if they were included in the same valence/activation category of the PANAS (e.g., the positive-valence, high-arousal items *celebrating* and *happy*). However, the variance decomposition (including persons, days, moments, and their interactions) showed largely consistent patterns across positive and negative moods, with variance at the person level accounting for the largest amount of variance across facets. This underscores the importance of considering person-level characteristics when measuring mood, for which some variance will usually be related to traits and/or stable environmental factors for each person. For example, comorbidities such as major depressive disorder and generalized anxiety disorder significantly impact mood and are more prevalent among people with OUD than in the general population; the potential presence of these disorders in our sample could be one explanation for the large proportion of mood explained at the person level. Consistently, the next largest proportion of variance was explained by the person-by-day interaction. This pattern was especially pronounced for negative mood items, suggesting that variance in negative mood for certain people may be particularly dependent on the type of day (e.g., weekdays versus weekends).

For people with addictions, days with structured vs. unstructured time (e.g., work vs. non-work days) have been shown to have protective (greater PA) and deleterious impacts on mood (greater NA) (Bertz et al., 2022; Epstein & Preston, 2012). This may be one, among potentially many, explanations for how the person-by-day variability in mood is unique for people with addiction. Interestingly, the facet capturing the person-by-moment interaction was smaller than the person and person-by-day facets—except for low-arousal mood states (e.g., *sleepy* and *fatigued*). This is likely because certain moments, for certain people, are more conducive to low arousal (likely just after waking up or just before going to bed). It is also important to note that a large proportion of variance was unexplained in our models. The highest percentage of unexplained variance was for the item *angry* (64%); the lowest, for *carefree* (28%). This unexplained variance not only includes measurement error, but also likely systematic influences from substantively meaningful sources not included in our models. Substantively meaningful sources may include aspects of persons, days, or moments known to influence mood; for example, a particular sensitivity to stressful events, the experience of multiple stressors within a day, and/or an experience of an acute stressor in a given moment, could independently and perhaps synergistically explain large proportions of variance in mood.

Relatedly, our finding that there was little or no variance within the main effect of moments does not suggest that variance is absent at moments in general; variance exists at the momentary level, but our results suggest it is unrelated to the ordering of our random EMA prompts.

To capture both the between- and within-person factor structure of mood in our OUD patients, we conducted a confirmatory multilevel factor analysis. Prior factor analyses have looked only at the between- and within-person factor structure of mood using daily data. These day-level factor analyses found a factor structure consistent with the PANAS—two factors, representing of positive and negative mood—at both the between- and within-person levels (Merz & Roesch, 2011; Rush & Hofer, 2014). However, it was important to test whether this factor structure would replicate at the within-person momentary level, because mood is a transient experience that is likely to be obscured if only assessed at the day level.

We found the best-fitting model to be one with three factors at the between-person and within-person momentary levels: positive, negative, and low-arousal mood factors. We further updated this model to capture the covariance between negative mood and low-arousal moods and the covariance of specific positive and negative mood items at the within-person level. Specifying low-arousal mood as a subfactor of negative mood slightly improved model fit, but was most consistent with a prior study showing that negative moods can include a low-arousal factor (Denollet & De Vries, 2006). The covariance between *angry* and *annoyed* for negative mood, and covariance between *pleased* and *contented* for positive mood, improved model fit, and also pointed to partial redundancy that could be reduced to ease participant burden in future EMA work. Although the reduction in item count (and associated burden) in this case would be relatively modest, such data-driven approaches could help rationalize questionnaire optimization more generally, including cases such as this where most items are shown to be non-redundant. Our approach comes with the caveat that our covariance specifications between these pairs of items were conducted

post hoc. Ideally, the findings from this data-driven approach would be specified *a priori* and supported by subsequent results to provide stronger support for our findings. Prior to arriving at this final model, we explored a factor structure more consistent with a circumplex model (PA and NA along with low and high arousal). The circumplex model was developed to depict valence and arousal as two bipolar factors, such that the states at either extreme (such as happiness and sadness, for valence) cannot be experienced simultaneously (Russell & Carroll, 1999). Subsequent evidence has generally pointed toward the existence of ambivalent states (Dejonckheere et al., 2021; Larsen & McGraw, 2011; Rafaeli & Revelle, 2006) and, in people with addictions, a connection between ambivalent states and craving or return to use (Menninga et al., 2011; Veilleux et al., 2013). As a result, we chose to test a circumplex-like structure of affect in terms of its four quadrants, eliminating the assumption that positive and negative moods cannot coexist: we treated each quadrant as unipolar. This resulted in the poorest fit at the between- and within-level for the SRMR, perhaps because arousal (unlike valence) might be better treated as a bipolar factor. Some authors have cautioned that factor analysis may not be able to extract bipolar factors from mood ratings even if bipolar factors are present (van Schuur & Kiers, 1994). To examine this, we also performed a series of diagnostic tests as outlined in prior investigations (e.g., van Schuur & Kiers, 1994) but did not find evidence in support of a bipolar factor structure.

Perhaps the most critical substantive finding from our study is the unique *three*-factor structure, with low-arousal moods constituting a separate factor (whereas high-arousal moods loaded onto the positive-valence or negative-valence factor). One possibility is that this factor structure is unique to our sample (people with addiction or OUD specifically): people diagnosed with addictive disorders tend to have higher rates of mood disorders than the general population, and they also commonly report reductions in positive moods, particularly during early abstinence from substance use (Garfield et al., 2014). This is important because low-arousal mood states can sometimes be difficult to disentangle from negative moods and reductions in positive moods; this is particularly common in the context of depression and especially anhedonia (Billones et al., 2020). In other clinical samples, some people with feelings of malaise or fatigue do not experience those feelings as distinctly negative or positive (Billones et al., 2020; Denollet & De Vries, 2006). Two prior MCFAs of daily mood, both conducted in non-clinical samples, each found only two factors (PA and NA) at the within- and between-person levels (Merz & Roesch, 2011; Rush & Hofer, 2014). This strengthens the possibility that our results may be indicative of a unique factor structure among (at least some) clinical samples. Our findings are also important because some measures currently used to assess mood (including the PANAS) include feelings of low arousal in a negative-mood factor. Our finding of a low-arousal factor suggests that low-arousal moods should be considered at least partly distinct from negative and positive moods. At the same time, our finding that the low-arousal factor can be treated as a subfactor of NA suggests that low-arousal states tended to be aversive, resembling NA states in that regard. Modifications to the PANAS, such as the PANAS-X (Watson & Clark, 1994), do offer a more comprehensive assessment of "other affective states," including low-arousal items similar to the ones we reported here. Greater adoption of the PANAS-X may be useful, perhaps particularly among clinical populations, though its length will need to be balanced against a potential increase in participant burden. In general, including a low-arousal factor

in routine assessment of mood in people with addictions may be important for understanding experiences in addiction recovery and predicting outcomes. Our research group has already conducted several investigations showing how mood states and traits are associated with important aspects of treatment outcome, arguably providing some signal of the concurrent validity of our proposed factor structure. Our prior findings include associations between: lower between-person PA and increased likelihood of treatment dropout; higher within-person NA and subsequent craving and, in some cases, substance use; and lower within-person sleep duration (perhaps a precursor to states of low arousal) and substance use (Bertz et al., 2019; Panlilio et al., 2019; Preston et al., 2018). However, our MCFA results can build on these findings by accounting for person- and momentary-level mood experiences in an integrated analysis. This measurement model will allow for follow up investigations with more precise tests of concurrent validity, simultaneously considering how different person- and momentary-level moods link to protective and deleterious treatment outcomes. For example, low-arousal states might subvert attending routine clinic visits and staying engaged with recovery supports; NA states might be more informative about risk of specific events like craving and substance use, and perhaps positive mood states and/or traits might serve as buffers against both low-arousal and NA states. Although these are speculations, it will be important to explore such possibilities to better understand PA, NA, and low-arousal states and their relative roles in addiction recovery.

Limitations and Future Directions

Our G-theory investigation showed a large proportion of variance unexplained by person, day, or moment, but this is as expected, given that we did not attempt to model the contributions of substantively meaningful drivers of mood (e.g., stressful stimuli, pleasant events, or encounters with drug-related cues). Delineating these drivers of mood with G-theory will be a particularly exciting area for future work. For example, understanding how substance-use events contribute to variance in mood at the person-, day-, and momentary-levels could inform the timing and frequency of interventions targeted to mitigate risky moods or enhance protective ones. Once-a-day mood interventions, for example, may be sufficient for changing mood at the day and momentary levels. Also, it is worth reiterating that our finding of little variance at certain levels (e.g., moments or days) does not suggest that variance is absent at moments or days in general; variance exists at these levels, but our results suggest it is unrelated to the pattern of assessment (e.g., ordering of prompts or days).

The first two models we fit came with the complexity of having a three-level factor structure (person, day, and moment). Though the three-level factor structure had poorer fit indices than other models, that may have been partly due to only a few moments nested within each day for each person (i.e., only the 3 random prompt questionnaires scheduled per day). This limitation is typical for EMA studies because more frequent assessment each day is burdensome, but the sparsity of moments within-day creates a challenge for fitting a three-level model. In our future studies, if more frequent moments are assessed within-day, it will come with reductions in the number of items to help minimize the possibility of response fatigue. Further, the distribution of items used in MCFA is important particularly when normality is assumed. Some mood items, particularly NA items, had skewed distributions, which may introduce bias in the parameter estimates of MCFA models, although that

bias is mitigated by estimation procedures in some software packages. The presence of non-normal input data is not synonymous with violations of normal distributional assumptions for multivariate factor loadings and residuals—the most pertinent assumptions of MCFA. Even so, one advantage of using visual analogue scales rather than Likert scales —as is increasingly common in EMA studies involving affect— is that continuous response items are generally more straightforward to analyze. An additional limitation of our study, though an important area for future investigation, is in understanding whether our factor structure is measurement-invariant across people and/or across time (McNeish et al., 2021). One way to address this might be to employ an EMA-burst design as part of a larger longitudinal study to assess mood before, during, and after addiction treatment. Testing differences in the factor structure of mood across these different EMA bursts could help determine whether the factor structure of mood is relatively fixed or dynamic in addiction. This investigation could include people with different types of SUDs (e.g., opioids, alcohol, and cannabis use disorder) to help determine the similarity or difference in factor structure of mood across different SUDs. It will also be necessary to test whether our three-factor solution replicates in people with no addictions. We suspect that if it does, it will be mostly for other clinical populations. In particular, it may be important to test our MCFA result among people with major depressive disorder, given the commonness of fatigue in depression and because aspects of reduced positive moods have some similarities between people with opioid addiction and people with depression (Denollet & De Vries, 2006; Stull et al., 2021). Next steps include criterion-validity tests with structural equation models to see how the factors resulting from our multilevel confirmatory factor analysis relate to important treatment outcomes in addiction. This may help clarify which mood factors (and perhaps even which items within factors) are most important for different outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Baker TB, Piper ME, McCarthy DE, Majeskie MR, & Fiore MC (2004). Addiction motivation reformulated: an affective processing model of negative reinforcement. Psychological review, 111(1), 33. [PubMed: 14756584]

Bertz JW, Epstein DH, Reamer D, Kowalczyk WJ, Phillips KA, Kennedy AP, Jobes ML, Ward G, Plitnick BA, Figueiro MG, Rea MS, & Preston KL (2019). Sleep reductions associated with illicit opioid use and clinic-hour changes during opioid agonist treatment for opioid dependence: Measurement by electronic diary and actigraphy. J Subst Abuse Treat, 106, 43–57. 10.1016/j.jsat.2019.08.011 [PubMed: 31540611]

Bertz JW, Panlilio LV, Stull SW, Smith KE, Reamer D, Holtyn AF, Toegel F, Kowalczyk WJ, Phillips KA, Epstein DH, Silverman K, & Preston KL (2022). Being at work improves stress, craving, and mood for people with opioid use disorder: Ecological momentary assessment during a randomized trial of experimental employment in a contingency-management-based therapeutic workplace. Behav Res Ther, 152, 104071. 10.1016/j.brat.2022.104071 [PubMed: 35390535]

Billones RR, Kumar S, & Saligan LN (2020). Disentangling fatigue from anhedonia: a scoping review. Translational Psychiatry, 10(1), 1–11. [PubMed: 32066695]

Burgess-Hull A, & Epstein DH (2021). Ambulatory Assessment Methods to Examine Momentary State-Based Predictors of Opioid Use Behaviors. Curr Addict Rep, 1–14. 10.1007/s40429-020-00351-7

Burgess-Hull AJ, Smith KE, Schriefer D, Panlilio LV, Epstein DH, & Preston KL (2021). Longitudinal patterns of momentary stress during outpatient opioid agonist treatment: A growth-mixture-model approach to classifying patients. Drug Alcohol Depend, 226, 108884. 10.1016/j.drugalcdep.2021.108884 [PubMed: 34229153]

Chen MK (1986). The epidemiology of self-perceived fatigue among adults. Preventive medicine, 15(1), 74–81. [PubMed: 3714661]

Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, & Bolger N (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? Personality and Social Psychology Bulletin, 32(7), 917–929. [PubMed: 16738025]

Dejonckheere E, Mestdagh M, Verdonck S, Lafit G, Ceulemans E, Bastian B, & Kalokerinos EK (2021). The relation between positive and negative affect becomes more negative in response to personally relevant events. Emotion, 21(2), 326–336. 10.1037/emo0000697 [PubMed: 31697104]

Denollet J, & De Vries J (2006). Positive and negative affect within the realm of depression, stress and fatigue: the two-factor distress model of the Global Mood Scale (GMS). J Affect Disord, 91(2-3), 171–180. 10.1016/j.jad.2005.12.044 [PubMed: 16458366]

Epstein DH, & Preston KL (2012). TGI Monday?: drug-dependent outpatients report lower stress and more happiness at work than elsewhere. The American journal on addictions, 21(3), 189–198. 10.1111/j.1521-0391.2012.00230.x [PubMed: 22494220]

Forgas JP (2008). Affect and Cognition. Perspect Psychol Sci, 3(2), 94–101. 10.1111/j.1745-6916.2008.00067.x [PubMed: 26158876]

Forrest SJ, Siegert RJ, Krägeloh CU, Landon J, & Medvedev ON (2021). Generalizability theory distinguishes between state and trait anxiety. Psychological assessment.

Furnari M, Epstein DH, Phillips KA, Jobes ML, Kowalczyk WJ, Vahabzadeh M, Lin JL, & Preston KL (2015). Some of the people, some of the time: field evidence for associations and dissociations between stress and drug use. Psychopharmacology (Berl), 232(19), 3529–3537. 10.1007/s00213-015-3998-7 [PubMed: 26153066]

Garfield JB, Lubman DI, & Yücel M (2014). Anhedonia in substance use disorders: a systematic review of its nature, course and clinical correlates. Australian & New Zealand Journal of Psychiatry, 48(1), 36–51. [PubMed: 24270310]

Grilli L, & Rampichini C (2007). Multilevel factor models for ordinal variables. Structural Equation Modeling, 14(1), 1–25. 10.1207/s15328007sem1401_1

Hoffman L, & Stawski RS (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. Research in human development, 6(2-3), 97–120.

Hox JJ, Moerbeek M, & Van de Schoot R (2017). Multilevel analysis: Techniques and applications. Routledge.

Hu L. t., & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. 10.1080/10705519909540118

Institute, S. (2014). SAS/STAT® 13.2 User's Guide.

Kline RB (2016). Principles and practice of structural equation modeling, 4th ed. Guilford Press.

Larsen JT, & McGraw AP (2011). Further evidence for mixed emotions. J Pers Soc Psychol, 100(6), 1095–1110. 10.1037/a0021846 [PubMed: 21219075]

McLellan AT, Luborsky L, Cacciola J, Griffith J, Evans F, Barr HL, & O'Brien CP (1985). New data from the Addiction Severity Index. Reliability and validity in three centers. J Nerv Ment Dis, 173(7), 412–423. 10.1097/00005053-198507000-00005 [PubMed: 4009158]

McNeish D, Mackinnon DP, Marsch LA, & Poldrack RA (2021). Measurement in Intensive Longitudinal Data. Structural Equation Modeling: A Multidisciplinary Journal, 1–16.

Medvedev ON, Krägeloh CU, Narayanan A, & Siegert RJ (2017). Measuring mindfulness: Applying generalizability theory to distinguish between state and trait. Mindfulness, 8(4), 1036–1046.

Menninga KM, Dijkstra A, & Gebhardt WA (2011). Mixed feelings: ambivalence as a predictor of relapse in ex-smokers. Br J Health Psychol, 16(3), 580–591. 10.1348/135910710x533219 [PubMed: 21722277]

Merz EL, & Roesch SC (2011). Modeling trait and state variation using multilevel factor analysis with PANAS daily diary data. Journal of Research in Personality, 45(1), 2–9. 10.1016/j.jrp.2010.11.003 [PubMed: 21516166]

Muthén BO (1991). Multilevel Factor Analysis of Class and Student Achievement Components. Journal of Educational Measurement, 28(4), 338–354. 10.1111/j.1745-3984.1991.tb00363.x

Panlilio LV, Stull SW, Bertz JW, Burgess-Hull AJ, Lanza ST, Curtis BL, Phillips KA, Epstein DH, & Preston KL (2021). Beyond abstinence and relapse II: momentary relationships between stress, craving, and lapse within clusters of patients with similar patterns of drug use. Psychopharmacology (Berl). 10.1007/s00213-021-05782-2

Panlilio LV, Stull SW, Kowalczyk WJ, Phillips KA, Schroeder JR, Bertz JW, Vahabzadeh M, Lin J-L, Mezghanni M, Nunes EV, Epstein DH, & Preston KL (2019). Stress, craving and mood as predictors of early dropout from opioid agonist therapy. Drug and alcohol dependence, 202, 200–208. [PubMed: 31357121]

Preacher KJ, Zyphur MJ, & Zhang Z (2010). A general multilevel SEM framework for assessing multilevel mediation. Psychol Methods, 15(3), 209–233. 10.1037/a0020141 [PubMed: 20822249]

Preston KL, Kowalczyk WJ, Phillips KA, Jobes ML, Vahabzadeh M, Lin JL, Mezghanni M, & Epstein DH (2017). Context and craving during stressful events in the daily lives of drug-dependent patients. Psychopharmacology (Berl), 234(17), 2631–2642. 10.1007/s00213-017-4663-0 [PubMed: 28593441]

Preston KL, Kowalczyk WJ, Phillips KA, Jobes ML, Vahabzadeh M, Lin JL, Mezghanni M, & Epstein DH (2018a). Before and after: craving, mood, and background stress in the hours surrounding drug use and stressful events in patients with opioid-use disorder. Psychopharmacology (Berl), 235(9), 2713–2723. 10.1007/s00213-018-4966-9 [PubMed: 29980821]

Preston KL, Schroeder JR, Kowalczyk WJ, Phillips KA, Jobes ML, Dwyer M, Vahabzadeh M, Lin JL, Mezghanni M, & Epstein DH (2018b). End-of-day reports of daily hassles and stress in men and women with opioid-use disorder: Relationship to momentary reports of opioid and cocaine use and stress. Drug Alcohol Depend, 193, 21–28. 10.1016/j.drugalcdep.2018.08.023 [PubMed: 30336389]

Rafaeli E, & Revelle W (2006). A premature consensus: are happiness and sadness truly opposite affects? Motivation and Emotion, 30(1), 1–12. 10.1007/s11031-006-9004-2

Rhemtulla M, Brosseau-Liard P, & Savalei V (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychol Methods, 17(3), 354–373. 10.1037/a0029315 [PubMed: 22799625]

Rush J, & Hofer SM (2014). Differences in within- and between-person factor structure of positive and negative affect: analysis of two intensive measurement studies using multilevel structural equation modeling. Psychol Assess, 26(2), 462–473. 10.1037/a0035666 [PubMed: 24512426]

Russell JA (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.

Russell JA, & Carroll JM (1999). On the bipolarity of positive and negative affect. Psychol Bull, 125(1), 3–30. 10.1037/0033-2909.125.1.3 [PubMed: 9990843]

Scott SB, Sliwinski MJ, Zawadzki M, Stawski RS, Kim J, Marcusson-Clavertz D, Lanza ST, Conroy DE, Buxton O, Almeida DM, & Smyth JM (2020). A Coordinated Analysis of Variance in Affect in Daily Life. Assessment, 27(8), 1683–1698. 10.1177/1073191118799460 [PubMed: 30198310]

Serby M (2003). Psychiatric resident conceptualizations of mood and affect within the mental status examination. Am J Psychiatry, 160(8), 1527–1529. 10.1176/appi.ajp.160.8.1527 [PubMed: 12900321]

Shacham S (1983). A shortened version of the Profile of Mood States. Journal of personality assessment.

Shavelson RJ, & Webb NM (1991). Generalizability theory: A primer. Sage Publications, Inc.

Shiffman S (2009). Ecological momentary assessment (EMA) in studies of substance use. Psychological assessment, 21(4), 486. [PubMed: 19947783]

Steptoe A, Wardle J, & Marmot M (2005). Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. Proc Natl Acad Sci U S A, 102(18), 6508–6512. 10.1073/pnas.0409174102 [PubMed: 15840727]

Stull SW, Bertz JW, Epstein DH, Bray BC, & Lanza ST (2021). Anhedonia and Substance Use Disorders by Type, Severity, and With Mental Health Disorders. Journal of Addiction Medicine.

Trull TJ, Solhan MB, Tragesser SL, Jahng S, Wood PK, Piasecki TM, & Watson D (2008). Affective instability: measuring a core feature of borderline personality disorder with ecological momentary assessment. J Abnorm Psychol, 117(3), 647–661. 10.1037/a0012532 [PubMed: 18729616]

van Schuur WH, & Kiers HAL (1994). Why Factor Analysis Often is the Incorrect Model for Analyzing Bipolar Concepts, and What Model to Use Instead. Applied Psychological Measurement, 18, 110–197.

Veilleux JC, Conrad M, & Kassel JD (2013). Cue-induced cigarette craving and mixed emotions: a role for positive affect in the craving process. Addict Behav, 38(4), 1881–1889. 10.1016/j.addbeh.2012.12.006 [PubMed: 23380484]

Watson D, & Clark LA (1994). The PANAS-X: Manual for the positive and negative affect schedule-expanded form http://www.psychology.uiowa.edu/Faculty/Watson/Watson.html

Watson D, Clark LA, & Tellegen A (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol, 54(6), 1063–1070. 10.1037//0022-3514.54.6.1063 [PubMed: 3397865]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Public Significance Statement

We found that moods in people receiving medication for opioid use disorder (MOUD) changed more between days than within days, and fell into three categories: positive, negative, and low-arousal (which partly overlapped with negative); high arousal did not form a separate category. These results could improve the way mood is assessed and reported, at least in MOUD patients, and provides a framework for examining dynamic qualities of mood in other populations.
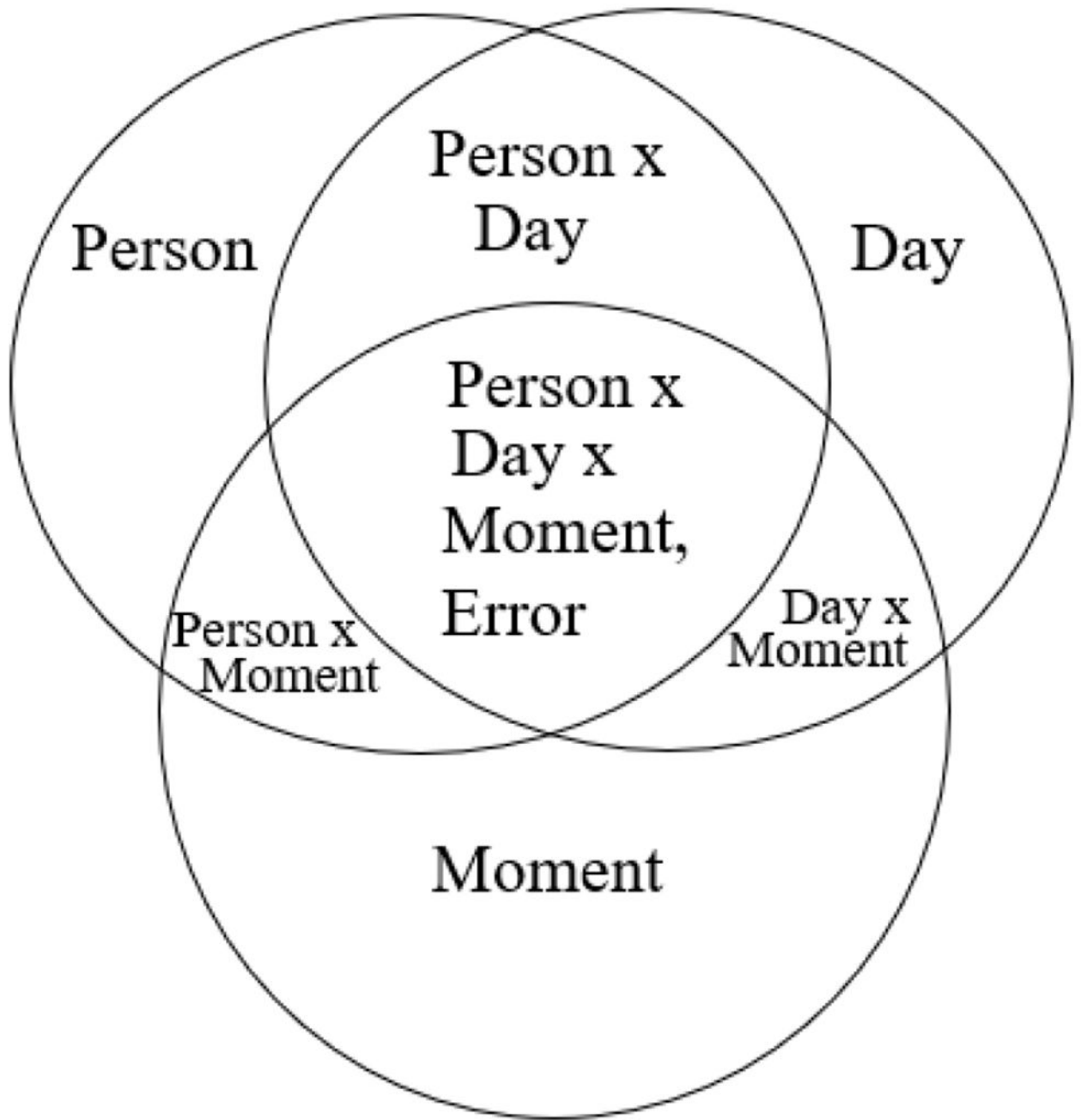
Figure 1.
Venn diagram of the mood-variance decomposition
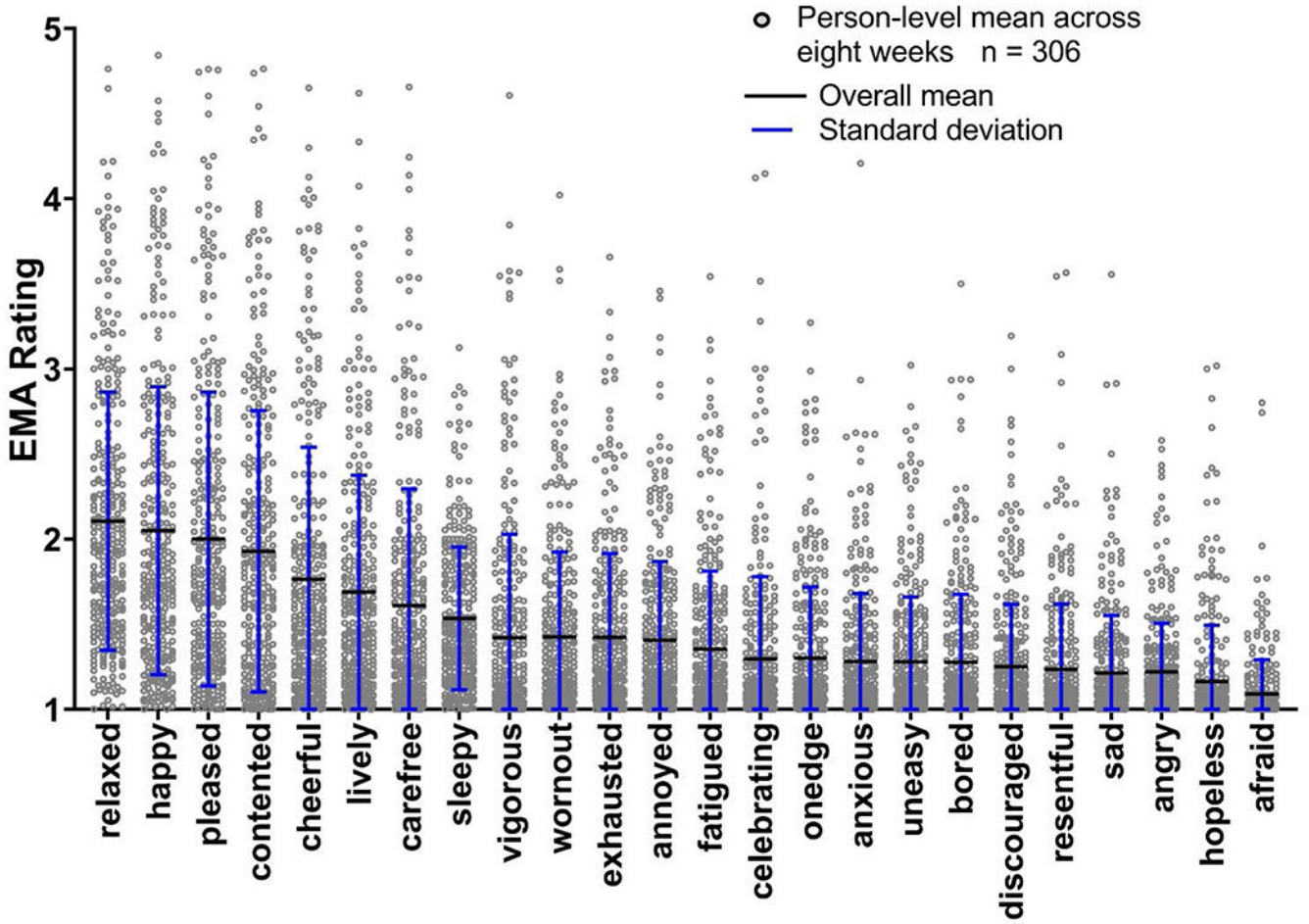
Figure 2.
Person-level means and overall means and standard deviations for EMA PANAS items.
Items ordered by descending overall mean.

Figure 3. Factor structure and standardized loadings for the final multilevel confirmatory factor analysis model

Note. All standardized factor loadings were statistically significant at p<.05; All factor correlations were statistically significant except for the correlation between PA and Low Energy at the person-level, as noted.

Table 1

Positive Mood Descriptive Statistics and Variance Partitioning

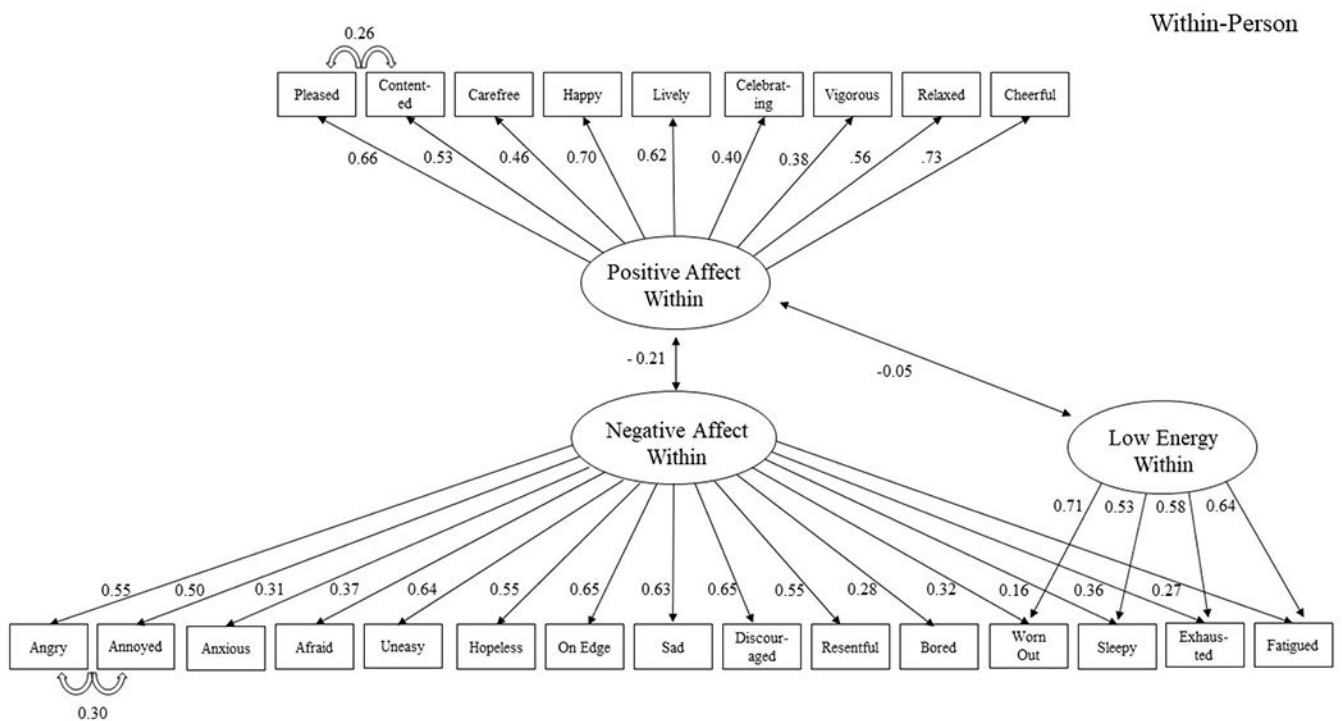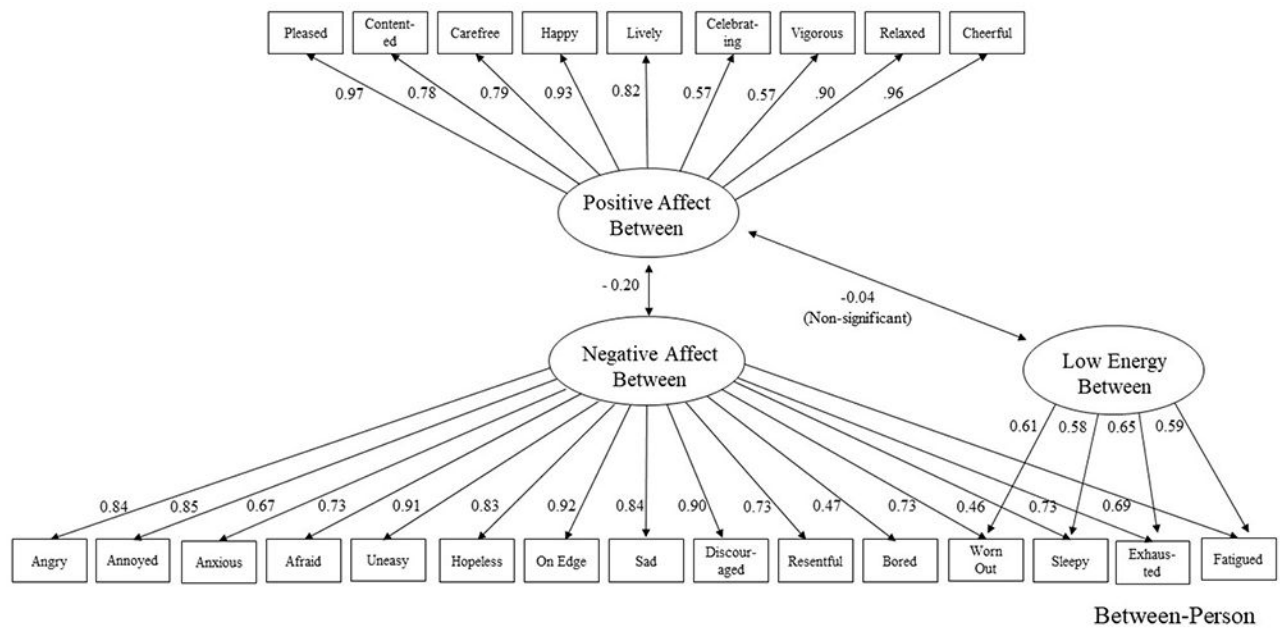| Positive Affect Item	Overall Mean (SD)	ICC	Overall Variance	% Variance:						
				Person	Day	Moment	Person*Day	Person*Moment	Day*Moment	Unexplained
Vigorous	1.42 (0.61)	0.57	0.68	60.9%	0.1%	0.0%	9.9%	0.0%	0.0%	28.9%
Pleased	2.00 (0.86)	0.58	1.31	60.2%	0.5%	0.0%	10.2%	0.1%	0.0%	28.9%
Contented	1.93 (0.83)	0.57	1.22	59.9%	0.1%	0.0%	11.1%	0.0%	0.0%	28.8%
Cheerful	1.76 (0.78)	0.56	1.09	59.2%	0.2%	0.0%	9.8%	0.4%	0.0%	30.3%
Carefree	1.61 (0.69)	0.56	0.86	58.9%	1.0%	0.0%	12.0%	0.0%	0.2%	28.0%
Happy	2.05 (0.85)	0.56	1.31	58.4%	0.4%	0.0%	12.0%	0.1%	0.0%	29.2%
Lively	1.69 (0.69)	0.50	0.94	51.5%	0.2%	0.1%	11.4%	0.8%	0.1%	35.6%
Relaxed	2.11 (0.76)	0.48	1.22	50.1%	0.1%	0.1%	11.2%	0.2%	0.0%	37.6%
Celebrating	1.30 (0.48)	0.46	0.51	49.2%	0.0%	0.0%	13.9%	0.0%	0.0%	36.8%

Note. ICC = Intraclass correlation coefficient

Table 2

Negative Mood Descriptive Statistics and Variance Partitioning

| Negative Affect Item | Overall Mean (SD) | ICC | Overall Variance | % Variance: | | | | | | | |
				Person	Day	Moment	Person*Day	Person*Moment	Day*Moment	Unexplained
Anxious	1.28 (0.40)	0.39	0.39	39.4%	0.1%	0.1%	11.9%	0.7%	0.1%	47.5%
On Edge	1.30 (0.42)	0.39	0.42	37.8%	0.0%	0.0%	12.7%	0.7%	0.0%	48.8%
Fatigued	1.35 (0.46)	0.41	0.47	37.7%	0.2%	0.1%	11.4%	1.7%	0.0%	48.9%
Worn Out	1.42 (0.50)	0.42	0.57	36.9%	0.2%	0.4%	13.1%	2.6%	0.0%	46.8%
Resentful	1.24 (0.38)	0.37	0.39	36.5%	0.0%	0.0%	15.0%	0.5%	0.0%	47.9%
Uneasy	1.28 (0.38)	0.36	0.40	35.9%	0.0%	0.0%	13.3%	0.2%	0.0%	50.5%
Bored	1.28 (0.40)	0.38	0.41	35.4%	0.1%	0.2%	11.2%	1.7%	0.0%	51.3%
Exhausted	1.42 (0.49)	0.38	0.60	29.2%	0.3%	0.4%	11.6%	2.5%	0.0%	50.3%
Hopeless	1.16 (0.33)	0.39	0.25	33.7%	0.0%	0.0%	16.3%	0.6%	0.0%	49.4%
Annoyed	1.41 (0.46)	0.34	0.59	31.7%	0.0%	0.0%	13.0%	0.3%	0.0%	55.0%
Discouraged	1.25 (0.37)	0.33	0.37	30.3%	0.0%	0.0%	15.4%	0.5%	0.0%	53.7%
Sad	1.21 (0.34)	0.31	0.36	29.2%	0.0%	0.0%	22.8%	0.3%	0.0%	47.7%
Afraid	1.09 (0.20)	0.27	0.15	27.9%	0.0%	0.0%	10.9%	0.0%	0.0%	61.1%
Sleepy	1.53 (0.42)	0.23	0.73	20.3%	0.7%	0.7%	10.8%	4.5%	0.0%	63.0%
Angry	1.22 (0.29)	0.21	0.36	17.4%	0.0%	0.0%	18.3%	0.5%	0.0%	63.8%

Note. ICC = Intraclass correlation coefficient

Table 3

Model fit statistics for multilevel confirmatory factor analysis models examining the PANAS

Model	Number of factors at each level	χ^2	df	p-value	TLI	RMSEA	SRMR Between	SRMR Within
Model 1 Levels: 3 (Person, day, and moment)	Person: 2 Day: 2 Moment: 2	16171.192	753	<.001	0.752	0.023	Person-level: 0.072 Day-level: 0.384	0.080
Model 2 Levels: 3 (Person, day, and moment) PA and NA item correlations specified at day-level	Person: 2 Moment: 2	13302.490	637	<.001	0.760	0.023	Person-level: 0.072 Day-level: 6.121	0.057
Model 3 Levels: 2 (Person, moment)	Person: 2 Moment: 2	16353.294	502	<.001	0.724	0.029	0.072	0.061
Model 4 Levels: 2 (Person, moment)	Person: 3 Moment: 3	7643.799	498	<.001	0.875	0.019	0.06	0.040
Model 4a Levels: 2 (Person, moment)	*Person: 3 * Moment: 3	7490.338	492	<.001	0.876	0.019	0.059	0.038
Model 4b Levels: 2 (Person, moment)	***Person: 3 *^ Moment: 3**	**5992.196**	**490**	**<.001**	**0.902**	**0.017**	**0.059**	**0.037**
Model 5 Levels: 2 (Person, moment)	Person: 4 Moment: 4	20318.774	492	<.001	0.648	0.032	0.095	0.065

Note. TLI = Tucker Lewis Fit Index, RMSEA = Root Mean Square Error of Approximation, SRMR = standardized root mean residual.

Bolded text to highlight Model 3b chosen as final model.

*
Low-arousal subfactor of NA (NA and Low Energy factors specified as uncorrelated)

^
Momentary -level residual variances specified between "angry" and "annoyed" and between "pleased" and "contented"