HHS Public Access
Author manuscript
Assessment. Author manuscript; available in PMC 2021 December 01.

Published in final edited form as:
Assessment. 2020 December ; 27(8): 1683–1698. doi:10.1177/1073191118799460.

A Coordinated Analysis of Variance in Affect in Daily Life

Stacey B. Scott,
Stony Brook University

Martin J. Sliwinski,
Pennsylvania State University

Matthew Zawadzki,
University of California-Merced

Robert S. Stawski
Oregon State University

Jinhyuk Kim, **David Marcusson-Clavertz**, **Stephanie T. Lanza**, **David E. Conroy**, **Orfeu Buxton**, **David M. Almeida**, **Joshua M. Smyth**
Pennsylvania State University

Abstract

Despite widespread interest in variance in affect, basic questions remain pertaining to the relative proportions of between- and within-person variance, the contribution of days and moments, and the reliability of these estimates. We addressed these questions by decomposing negative affect (NA) and positive affect (PA) variance across three levels (person, day, moment), and calculating reliability using a coordinated analysis of seven daily diary, ecological momentary assessment (EMA), and diary-EMA hybrid studies (across studies age = 18–84 years, total N_{persons} = 2,103, total $N_{\text{observations}}$ = 45,065). Across studies, within-person variance was sizeable (NA: 45–66%, PA: 25–74%); in EMA more within-person variance was attributable to momentary rather than daily level. Reliability was adequate to high at all levels of analysis (within-person: .73-.91; between- person: .96–1.00) despite different items and designs. We discuss the implications of these results for the design of future intensive studies of affect variance.

Correspondence should be addressed to: Stacey Scott, Department of Psychology, Stony Brook University, Stony Brook, NY 11793 stacey.scott@stonybrook.edu or Martin Sliwinski, Center for Healthy Aging, Pennsylvania State University, University Park, PA 16802 mjs56@psu.edu.
Author Note
Stacey B. Scott, Department of Psychology, Stony Brook University.
Martin J. Sliwinski, Department of Human Developmental & Family Studies, Center for Healthy Aging, Pennsylvania State University.
Matthew Zawadzki, Department of Psychological Sciences, University of California-Merced.
Robert S. Stawski, School of Social and Behavioral Health Sciences, Oregon State University.
Jinhyuk Kim, Center for Healthy Aging, Pennsylvania State University.
David Marcusson-Clavertz & Stephanie T. Lanza, Department of Biobehavioral Health, Pennsylvania State University.
David E. Conroy, Department of Kinesiology, Pennsylvania State University.
Orfeu Buxton, Department of Biobehavioral Health, Pennsylvania State University.
David M. Almeida, Department of Human Development & Family Studies, Center for Healthy Aging, Pennsylvania State University.
Joshua M. Smyth, Department of Biobehavioral Health, Pennsylvania State University.

Keywords

affect; variability; ecological momentary assessment; daily diary; intensive longitudinal design

Fluctuations in affective states are a focus in research relevant to a broad range of outcomes including well-being in aging (Carstensen et al., 2011), personality (Eid & Diener, 1999; Kuppens, Van Mechelen, Nezlek, Dossche, & Timmermans, 2007), psychopathology (Ebner-Priemer, Eid, Kleindienst, Stabenow, & Trull, 2009; Gross & Jazaieri, 2014) and physical health symptoms (Pressman & Cohen, 2005; Watson, 1988). Research on affect variance typically uses intensive longitudinal designs (ILD), including ecological momentary assessment (EMA) studies that prompt individuals multiple times per day to ask about current or recent states, daily diary surveys that probe about the day overall, or hybrid designs involving both momentary reports throughout the day and an end of day (EOD) retrospective survey of the day overall. These different designs enable an investigation of whether an individual's affect shows meaningful (i.e., reliable, predictable) variance over different time scales. With few exceptions, however, researchers have examined "within-person" affect variance without distinguishing variance observed across the momentary (within-day) and daily (across-day) time scales. As a consequence, fundamental questions such as "does affect vary more within a day than across days?" and "do momentary and daily changes in affect reflect reliable variation?" remain unanswered. These questions are important because numerous studies have used indices of within-person affect variance to characterize developmental processes, adaptation, and risk for health outcomes (Bisconti, Bergeman, & Boker, 2004; Brose, Scheibe, & Schmiedek, 2013; Watson, 1988).

This manuscript addresses three interrelated aims regarding the assessment of affect variance, with an overall goal to inform the design of future studies. Our first aim was to answer the question of time scale of affective variance. For both negative affect (NA) and positive affect (PA), we decomposed the variance to examine the relative amount of between- and within-person variance; for EMA data, we further decomposed within-person variance into variance across moments versus across days. Our second aim was to investigate whether the within-person affect variance would be replicated across studies of different populations using different instruments. To do so, we decomposed the percent variance in an observation due to between-person differences, measurement occasions, items and their interactions. Our third aim was to inform future ILD studies by comparing reliability estimates across different numbers of items. Results from any single study can reflect the peculiarities of that study's design, sample, affect items, and response categories. One possible approach, integrative analysis (i.e., pooling data from multiple studies and analyzing as a single dataset; Curran & Husson, 2009), was not appropriate because of the heterogeneity in methods across studies. Instead, we conducted coordinated analyses (Hofer & Piccinin, 2009) to directly study the comparability and replication of findings from multiple studies of varying populations which used different instruments to assess the same constructs. To this end, we applied the same analytic models to data from seven different ILD studies that measured NA and PA — three EMA studies, three daily diary studies and two hybrid studies that used both EMA and diary-style EOD reports. This approach strengthens conclusions through replication, a hallmark of rigorous experimental work despite being relatively uncommon in naturalistic studies.

Affective Variance in Daily Life: Reliably Assessing Persons, Days, and Moments

Within-person affective variance refers to fluctuations in emotional states within individuals across time or situations. There are two ways that researchers generally index affect variance (see Ram & Gerstorf, 2009): stochastic and systematic variation. Stochastic variance refers to volatility, unpredictability, or instability in levels of affect. Examples include studies which the total amount of affect variance is quantified by some index such as the intraindividual standard deviation, the root mean square successive difference, or entropy (Moskowitz & Zuroff, 2004). In contrast, systematic variance refers to time-ordered variations in a person's internal states or external environment. Examples include regression coefficients that reflect constructs such as emotional inertia or stressor reactivity (Suls, Green, & Hillis, 1998). Both approaches assume that the magnitude of the differences among repeated observations within an individual conveys useful information (i.e., signal) about that person.

There is seldom strong justification as to why within-person variance is examined across a particular time scale (e.g., moments, days). How much an individual's affect varies across moments within a day could reflect the structure of her or his day, diurnal variations in arousal, or the transient effects of the individual's current state of mind or surroundings. Day-to-day variations could reflect workday/non-workday status or the effects of a bad night's sleep or an impending deadline. In addition to the contextual explanations for what is occurring in a moment versus what was the nature of the day, there are measurement explanations for why moment- and day-level variance are not interchangeable (Gorin & Stone, 2001). Daily diary and other studies using EOD assessments often frame questions by asking participants to retrospectively rate the day overall, whereas EMA studies often instruct participants to rate their current state (or over a much shorter recent interval). Despite the increasingly common use of ILD approaches, however, it is still unclear whether variance across days is unique or simply a by-product of variation on a faster time scale (momentary). One focus of this manuscript is to characterize the reliable within-person variance across momentary and daily timescales, different assessment approaches (EMA, EOD), and different sets of affect items.

Despite the widespread interest in affective variance, the unresolved issues of the time scale and reliability of these within-person fluctuations hampers future advances. ILD studies on affect differ greatly in the frequency and duration of assessments, items (e.g., number, content, phrasing, response options) queried at each assessment, and sample characteristics. Studies also differ in the analytic approaches - with some researchers reporting individual standard deviations (iSD) and others percentages of between- vs. within-person variance based on intraclass correlation coefficients (ICC). EMA data are variously analyzed using, for example, either two levels (moments within people) or three levels (moments within days within people). Thus, it is difficult for researchers designing studies and planning analyses to glean from the literature whether individuals vary in their affect more within days than across days. Similarly, in line with the relative dearth of careful psychometric information available for ILD, there is little information regarding the number of items needed to reliably assess PA or NA in ILD studies. Quantifying affective variance assumes that short-term changes in affect across days or across moments within a day can be reliably measured, and that such changes reflect variation across that specific time scale. A few previous studies have addressed this topic, but have generated some seemingly contradictory results. For example, some studies have demonstrated reliable within-person variance at the daily time scale (Cranford et al., 2006), whereas others have not (de Haan-Rietdijk, Kuppens, & Hamaker, 2016).

In sum, this study addresses three aims: (1) decompose variance in PA and NA due to between- person differences and within-person fluctuations, and with EMA data further describe the proportion of within-person variance at the day and momentary level, (2) estimate between-person reliability (e.g., reliability of a measure of stable between-person differences in affect) and estimate within-person reliability (i.e., reliability of a measure of differences between occasions within the same person), and (3) provide recommendations about the number of items needed to construct reliable measures of between-person differences and within-person fluctuations across occasions in PA and NA.

Method

These coordinated analyses utilized data from ILD studies: two daily diary studies, three EMA studies, and two hybrid designs in which both EMA-style momentary and daily diary-style end of day reports were collected (see Smyth et al., in press). We present brief descriptions of the samples and procedures below. Affect items and assessments modes are presented in Tables 1 and 2 in order to allow for side-by-side comparison. For brevity, descriptions of hybrid studies in Tables 1 and 2 are organized such that the momentary reports are described with the EMA studies and the end of day reports are described with the daily diary studies. Data collection for all studies was approved by their respective Institutional Review Boards; the coordinated analysis of secondary data was deemed exempt by the Institutional Review Board at Pennsylvania State University.

Daily Diary Studies: End of Day Reports.

Daily Study I: National Study of Daily Experiences (NSDE).—Data were drawn from the second wave of the National Study of Daily Experiences (NSDE), a part of the larger Midlife in the United States study (MIDUS). For additional detail on the sample and study protocol, see Almeida, McGonagle, and King (2009).

Participants.: Daily diary data was collected from 2,022 individuals who previously participated in the larger Midlife in the United States study (MIDUS; *N*=4,963). Of the 2,022 NSDE respondents, 1,079 were from the random digit dialing (RDD) sample, 185 siblings of individuals in the RDD sample, 516 from the twin RDD subsample, 62 from the city oversamples, and 180 from the Milwaukee-specific subsample. As multiple individuals from families are represented, we randomly selected a single family member from each family for the current analyses; resulting in 1,689 participants. The reliability calculation could not be conducted due to insufficient memory with this full sample; thus, we used a simple random sample of half the participants and used this for all analyses reported[1]. The resulting analysis sample was comprised of 845 participants with an average age of 56.3 (*SD*=12.2, Range=33–84), 58% were female, 38.3% had received a high school diploma or less, 26.8% has completed some college, and 35.0% had completed a bachelor's degree or more.

Procedure.: Participants completed telephone interviews (~20 minutes) on 8 consecutive evenings. Each 8-day interview protocol consisted of separate random subgroups of 30 participants with the start day staggered across the day of the week to control for the possible confounding between day of study and day of week. Based on the number of participants (N = 845) and study days (n = 8), the maximum number of daily observations possible was 6,760; 6,198 (91.78%) daily observations were collected. NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 1 for each table).

Daily Study II: Work, Family, & Health Study (WFHS).—The Work, Family and Health Study is a multisite workplace intervention conducted in both information technology and extended care (e.g., nursing home) workplace contexts (Bray et al., 2013). As part of the larger intervention study, a daily diary study was conducted using a subsample of study participants. For the purposes of the current study we used data from the first measurement burst of daily diary assessments.

Participants.: The sample comprised 313 adults with a mean age of 41.4 (SD = 7.11, Range = 21–63), 74% were female. In terms of education, 3.5% had some high school, 18.0% were high school graduates, 39.9% had some college or technical trainings, and 38.6% had 4 or more years of college.

Procedure.: Participants completed telephone interviews on 8 consecutive evenings following a protocol similar to that of NSDE described above. Based on the number of participants (N = 313) and study days (n = 8), the maximum number of daily observations possible was 2,504; 2,311 (92.3%) daily observations were collected. NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 2 for each table).

Ecological Momentary Assessment Studies: Momentary Reports.

EMA Study I: Stress, Health, and Daily Experiences (SHADE).—The Stress, Health, and Daily Experiences (SHADE) study sought to examine how daily experiences relate to health and well-being among people with chronic disease, namely asthma and rheumatoid arthritis (RA). For additional details on the sample and study protocol, see Smyth, Zawadzki, Santuzzi, & Filipkowski (2014).

Participants.: Participants were recruited via print media and television and radio advertisements. Participants (*N* = 128) met with a physician to confirm diagnosis of RA (*N*

[1]As a fidelity check, we did a follow-up in which we created four simple random samples from the NSDE dataset and conducted parallel analyses on each. The pattern of results was consistent across these subsamples, thus for simplicity here and in the macro available on our project site, we present the results based on one simple random sample.

= 97) or asthma (*N* = 31). Of the total sample, 117 (91%) participants provided EMA data. Participants had a mean age of 44.2 (*SD* = 14.2, Range = 18–80) and were predominantly Caucasian (84%) and female (73%). Exclusion criteria consisted of the following: younger than 18 years of age; no clinically verified diagnosis of RA or asthma; current drug or alcohol abuse problems; receiving emergency room treatment (other than minor injury), having a medication or other treatment change, or receiving a diagnosis of a mental illness within the prior three months; or being unable to complete the EMA protocol (e.g., due to poor eyesight).

Procedure.: After being screened for eligibility, participants came to the laboratory and completed baseline measurements not relevant to the present study. They were then trained on how to use a provided palmtop computer. The palmtop computers were programmed to alert participants to complete surveys 5 quasi-random times each day for 7 days. With 117 participants, 7 days, and 5 momentary assessments daily, the maximum number of momentary observations possible was 4,095; 3,553 (86.8%) momentary observations were collected. NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 5 for each table).

EMA Study II: North Texas Heart (NTH).—The North Texas Heart (NTH) study sought to examine social vigilance as a predictor of cardiovascular disease. For additional details on the sample and study protocol, see Ruiz and colleagues (In Press).

Participants.: A diverse community sample from the North Texas area was recruited through advertisements in local newspapers, flyers, community and university websites, and hospital postings. Participants (*N* = 300) were sampled as stratified by gender within age and race/ethnicity resulting in the following demographics: 150 men, 150 women; aged 21–70 (*M* = 42.44, *SD* = 12.76); 60% non-Hispanic Whites, 15% non-Hispanic Blacks, and 19% Hispanic/Latino/a. Exclusion criteria consisted of the following: unable to give informed consent, having a previous history of myocardial infarction, pregnancy within the past 12 months, and being a night shift worker.

Procedure.: After being screened for eligibility, participants arrived at a community vascular medicine clinic on a Thursday morning. They provided consent, underwent a brief physical exam, completed a personal and family medical history, gave a fasting blood draw, and completed a battery of surveys. Finally, prior to leaving all participants were fitted with an ambulatory blood pressure monitor and given a cellular phone to complete the EMA protocol. For two consecutive days, participants completed the EMA roughly every 45 minutes during waking hours; an EMA report was completed after each blood pressure measurement (programmed to occur at random times within 45-minute intervals). Because of different start times to the study, wake and sleep times, and blood pressure functions, participants varied in the number of observations they completed. A total of 8,136 observations were collected. NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 6 for each table).

EMA Study II: Work & Daily Life (WDL).—The Work and Daily Life (WDL) study examined how workplace stress affects health and well-being among a sample of full-time employed adults. For additional details on the sample and study protocol, see Damaske, Smyth, & Zawadzki (2014).

Participants.: Participants (*N* = 122) from the greater metropolitan area of a mid-sized city in the Northeast were recruited for a study measuring work characteristics and health. Participants had a mean age of 41.2 (SD = 11.62, Range = 19–63) and were predominantly Caucasian (76.1%) and female (74.5%). Exclusion criteria consisted of the following: younger than 18 years of age; not currently employed Monday through Friday with regular working hours between 6:00am and 7:00pm; employed on weekends; unable to come to the research laboratory on a Wednesday evening and the following Monday; not fluent in English; pregnant; and having a psychiatric therapy or drug treatment change within the prior three months. Of the total sample, 115 (94.3%) participants provided EMA data.

Procedure.: Participants were recruited via random calls from a local telephone directory and from public listings on a university e-mail news alert and local event websites. After being screened for eligibility, participants were trained on how to use a provided palmtop computer. Palmtop computers signaled participants to complete momentary surveys 6 quasi-random times each day for 3 days (Thursday-Saturday). With 115 participants, 3 days and 6 momentary assessments per day, the maximum number of observations possible was 2,070; 1,852 (89.5%) were collected. NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 7 for each table).

Hybrid Studies: Momentary and End of Day Reports.

Hybrid Study I: Effects of Stress on Cognitive Aging, Physiology, and Emotions (ESCAPE).—Data were drawn from the first measurement burst of the longitudinal Effects of Stress on Cognitive Aging, Physiology, and Emotion (ESCAPE) study. For additional detail on the sample and study protocol, see Scott and colleagues (2015).

Participants.: Participants were recruited via letters and phone calls using systematic probability sampling of New York City Registered Voter Lists for the zip code 10475, an area of Bronx, NY. Eligibility criteria included between 25 and 65 years of age, ambulatory, fluent in English, without visual impairment, and a resident of Bronx County. Participants (*N* = 242) ranged in age from 25 to 65 years (*M* = 46.77, *SD* = 10.88); women made up 66.39% of the sample. Sample sizes differed slightly between EMA (*N* = 241) and EOD (*N* = 240): two participants completed EMA but not EOD, one participant completed EOD but not EMA. The sample was diverse in terms of racial and ethnic identity: 9.13% identified as Non-Hispanic White, 63.07% as Non-Hispanic Black, 17.84% as Hispanic White, 5.81% as Hispanic Black, 0.41% as Asian, and 3.73% as Other.

Procedure.: As part of the larger study, participants visited the research offices and received training on the use of study smartphones to complete the affect surveys. Participants carried the specially-programmed study smartphones for 14 days. *Momentary data collection*: The smartphones beeped 5 times each day during the 14-day study period to prompt participants to complete momentary surveys; smartphones were programmed to beep based on participants' self-reported typical waking time. The average time between scheduled beeps was 2 hours and 33 minutes. Based on the number of participants (N = 241), days (n = 14), and momentary assessments (n = 5 daily), the maximum number of momentary observations would be 16,870; 13,966 momentary observations were collected. Momentary NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 8 for each table). *End of day data collection*: Prior to bedtime each day, participants self-initiated a daily diary survey on the smartphone. At the end of the 14 days, participants returned phones to the lab and completed additional assessments. Based on the number of participants (N = 240) and study days (n = 14), the maximum number of momentary observations possible was 3,360; 2,753 (81.9%) daily observations were collected. EOD NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 3 for each table).

Hybrid Study II: Stress and Working Memory (SAWM).—Data were drawn from the first measurement burst of the longitudinal Stress and Working Memory (SAWM) study (Mogle, Muñoz, Hill, Smyth, & Sliwinski, 2017).

Participants.: Participants were recruited from advertisements and flyers in a city in the Northeast U.S. Eligibility criteria included being between 20 to 80 years of age, able to operate a palm-top computer, and lack of major cognitive impairment. Participants (N = 174) ranged in age from 20 to 79 years (M = 49.45, SD = 16.90); women made up 51.14% of the sample. Sample sizes differed slightly between EMA (N = 172) and EOD (N = 170): four participants completed EMA but not EOD, two participants completed EOD but not EMA. More than half (57.89%) of the sample identified as Non-Hispanic White, 31.58% as Non-Hispanic Black, 3.51% as Hispanic Black, and 7.02% as other race or ethnicity.

Procedure.: As part of the larger SAWM study, participants attended a lab training session on the protocol and how to operate the palm-top computers to complete affect surveys. Participants carried the palm-pilots for 7 days and returned the equipment to the lab at the end of this period to complete additional study tasks. *Momentary data collection*: Palm-top computers were programmed to beep 5 quasi-random times daily based on participants' self-reported wake time. Participants were instructed to complete a momentary survey after each beep. The average time between beeps was 2 hours and 41 minutes. Based on the number of participants (N = 172), days (n = 7), and momentary assessments (n = 5 daily), the maximum number of momentary observations possible was 6,020; 5,239 (87.0%) momentary observations were collected. Momentary NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 9 for each table). *End of day data collection*: Before bed each night, participants completed a self-initiated daily diary survey on the palm-pilot. Based on the number of participants (N = 170) and study days (n = 7), the maximum number of daily observations possible was 1,900; 1,061 (89.2%) daily observations were collected. Momentary NA items are listed in Table 1, PA items in Table 2, and assessment modes in Table 3 (column 4 for each table).

Analysis.

For each study, NA and PA scores were calculated at each assessment for each scale by averaging across items. Our approach to coordinated analysis was to conduct separate but parallel analyses across each dataset. We provide a SAS macro these analyses at our Open Science site https://osf.io/y3mfe/ and https://osf.io/p4kwm/.

Decomposing Variance.—For our first goal of decomposing variance in NA and PA, we conducted unconditional multilevel models separately for NA and PA in each EOD and momentary dataset. Two-level models which partitioned variance into between-person (i.e., variance due to differences between individuals) and within-person (i.e., variance due to fluctuations within individuals across occasions and error[2]) levels were used for datasets with EOD reports. Three-level models were used for datasets with momentary reports. These models partitioned variance into the proportion due to differences between individuals (i.e., between-person variance) and two within-person levels: variance within individuals across days (i.e., within-person, across-day variance) and variance within individuals within days (i.e., within-persons, across moments and error). We then examined the variance components for each model and calculated the percent of total variance at each level (i.e., person, day, and moment)[3].

Calculating Reliability.—As above, we conducted separate but parallel analyses in each dataset in order to examine calculate the generalizability coefficient for the variance due to person, occasion, item, and the interactions of personXitem, occasionXitem, personXoccasion, and error using the VARCOMP procedure in SAS. Error refers to the residual variance not accounted for by the other design features in this decomposition (i.e., personXoccasionXitem). We followed Cranford and colleagues' (2006) equation[4] for R_c, which describes the precision of measurement to detect systematic change within-persons, an estimate of within-person reliability. Their calculation used EOD reports to describe the reliability of a daily NA measurement (and in a separate model, PA) for detecting systematic change in mood from day to day. In our equations, p = person, k = occasion (e.g., moment or day), e = error, i = item, m = number of items. Here, we extended this model to accommodate both daily and momentary data from EMA, diary, and hybrid designs. We refer to it as R_{WP}:

$$R_{WP} = \frac{\sigma^2_{p*k}}{\sigma^2_{p*k} + \left[\frac{\sigma^2_e}{m}\right]}$$

[2]In these models, the lowest level within-person variance (i.e., day in EOD, moment in EMA) is given by the residual variance in the covariance matrix. This residual term includes "error," which is operationalized as inter-item variance.
[3]de Haan-Rietdijk and colleagues (2016) directed researchers not to rely on the significant tests from these empty models when determining whether to use a three- or two-level structure; instead, they recommend using AIC to examine the results of autoregressive multilevel models. This approach is important for determining whether higher order variance structures are needed once lower ordered variance is taken into account. As our research questions are about estimating the relative contributions of variance at different levels and detecting time scales over which reliable variance can be measured, we do not take this alternate approach.
[4]Nezlek (2016) reviews the merits and weaknesses of various approaches for calculating within-person reliability in ILD and recommends the Cranford and colleagues approach for fixed daily diary designs.

R_{WP} describes how reliable each dataset's NA or PA scale was for detecting systematic change in mood from moment to moment in EMA or hybrid momentary data and day to day in daily diary or hybrid EOD data, reflecting the within-person reliability. The residual error variance in this equation is divided by m items to account for the precision gained by averaging over the fixed set of m items.

In addition to allowing examination of within-person variance, ILD studies can provide highly reliable estimates of person-level averages. Thus, researchers may use ILD approaches to generate and examine stable or semi-stable individual differences in affect. In order to estimate the reliability of person-level averages of observations across the entire study, we used Cranford and colleagues' equation[5] for R_{KF}. In this equation, the residual error variance is divided by km to account for the fixed number of days and items over which the person's average score is computed. For clarity, we refer to this estimate of reliability for detecting between-person differences as R_{BP}.

$$R_{BP} = \frac{\sigma_p^2 + \left[\frac{\sigma_{p*i}^2}{m}\right]}{\sigma_p^2 + \left[\frac{\sigma_{p*i}^2}{m}\right] + \left[\frac{\sigma_e^2}{km}\right]}$$

Solving for Number of Items.—Our final calculations build upon the results of the above analyses in order to answer questions regarding how many items are needed to construct a reliable measure of differences between occasions within the same person (R_{WP}) and stable differences between individuals (R_{BP}) to achieve a particular level of reliability. We followed the examples provided by Segerstrom and colleagues (2014). For each study, we used the VARCOMP results for p, i, e, and k and then solved for R_{WP} and R_{BP} with m items ranging from two to eight.

Results

Affect Variance Decomposition: Persons, Days, & Moments.

Results for the decomposition of NA variance are displayed in Figure 1; between-person portion variance in each dataset is shown in the grey shaded area of the bar, within-person variance is displayed in the unshaded area. As described above, in EMA datasets within-person variance can be due to two sources of reliable variance: within-persons across days (the lower portion of the unshaded area) and variation within-persons across moments within days (the upper dotted portion of the unshaded area). Across EOD datasets with different numbers of items and indicators of NA, 43%−55% (Mean = 50%, Median = 52%) of the NA variance was due to differences between individuals. This implies that 45%−57% (Mean =

[5]Cranford and colleagues used the maximum number of study days to calculate this in their daily diary studies. Although all the studies employed in our analysis had good compliance, it is rare for a majority of participants to complete every possible daily diary phone interview, end of day survey, or momentary beeped prompt. In order to account for this, we calculated between-person reliability and within-person reliability for each study based on the number of observations completed on average (i.e., average compliance X maximum observations possible); this provides a more conservative estimate. We also calculated using maximum observations possible but do not present these in the tables nor text. Estimates using average observations completed compared to maximum possible differed by <.003.

50%, Median = 48%) of the NA variance in these datasets was due to fluctuations within-persons (including both systematic and stochastic). Across momentary datasets with different items and response scale options, 34%−52% (Mean = 44%, Median = 42%) of the NA variance was due to between-person differences. The remaining variance was partitioned into within-person variance at the level of days and moments. Within-person across day variance in NA ranged from 5%−19% (Mean = 13%, Median = 15%), whereas within-person across moment variance, including both systematic and stochastic, ranged from 34%−61% (Mean = 43%, Median = 40%) of the total NA variance.

Decomposition of PA variance is displayed in Figure 2. In EOD datasets, 57%−75% (Mean = 63%, Median = 60%) of the variance was due to between-person differences, implying that 25%−43% (Mean = 37%, Median = 40%) of PA variance was due to fluctuations within-persons (including both systematic and stochastic). In EMA datasets, 26%−56% (Mean = 44%, Median = 50%) of PA variance was at the between-person level, whereas the remaining within-person variance was decomposed into 7%−22% (Mean = 14%, Median = 13%) due to within-person across day and 31%−67% (Mean = 42%, median = 37%) due to within-person across-moment variance (including systematic and stochastic).

It is important to note here that the above results do not imply that there is systematically less within-person variance in PA compared to NA, rather that the ratios of between-person to within-person variance differ across PA and NA. In general, PA had larger variance estimates compared to NA. Figure S1 displays the raw variance across the studies.

Reliability to Detect Individual Differences and Within-person Fluctuations.

Within-person Reliability.—Within-person reliability estimates for NA and PA for EOD studies are displayed in Table 4. In EOD datasets, R_{WP} for NA ranged from .73-.85 (Mean = .80, Median = .81). PA R_{WP} ranged from .81-.89 (Mean = .85, Median = .84) in EOD datasets. Reliability estimates for EMA datasets are displayed in Table 5. In EMA datasets, R_{WP} for NA ranged from .78-.85 (Mean = .82, Median = .83) and for PA estimates ranged from .78-.91 (Mean = .85, Median = .85).

Between-person Reliability.—Between-person reliability estimates for NA and PA for each study are displayed in Table 4 for EOD and Table 5 for EMA. In EOD datasets, NA R_{BP} ranged from .95-.99 (Mean = .97, Median = .97); PA R_{BP} ranged from .98-.99 (Mean = .99, Median = .99). In EMA datasets, NA R_{BP} ranged from .99 to 1.00 (Mean = .99, Median = .99); PA R_{BP} ranged from .99 to 1.00 (Mean = 1.00, Median = 1.00).

Number of Items Required to Achieve Desired Reliability.

Within-person Reliability.—Figure 3 displays the within-person reliability estimates for different numbers of items across our studies; Tables S2-S5 show the estimates for Figure 3. In EOD assessments, our median estimates for adequate within-person reliability of .6 (i.e., adequate reliability) are 3 NA and PA items. For high within-person reliability of .8 (i.e., high reliability) in EOD studies, our median estimates are 9 NA and 6 PA items. For an adequate within-person reliability in EMA studies, we estimate that a median of 2 NA and

PA items would be needed. For high within-person reliability, our median estimates are 4 NA and 3 PA items in an EMA study.

Between-Person Reliability.—Given the high between-person reliability, values were at ceiling across 2 or more items. The estimates, however, are provided in Tables S2-S5. For both EOD and EMA studies, the median estimates indicate that high between-person reliability was achieved with 2 items for NA and PA.

Discussion

Given the emergence of affect variance as an indicator of emotional lability associated with or reflective of constructs such as stress reactivity, we used a coordinated analysis approach to examine between- and within-person variance in NA and PA in seven studies of self-reported affect in everyday life. We also calculated between- and within-person reliability of these assessments. These studies varied in their samples, items, response formats, and measurement frequencies but each asked participants to report on their negative and positive affective experiences in daily life.

Persons, Days, and Moments

Across EOD and EMA studies the median results indicate that there is about a 50:50 split of between to within-person variance in NA, whereas in PA we observed somewhat more between-person variance with about 44% at the within-person level. The figures show remarkable consistency across the studies, with exception of relatively low between-person NA variance in one EMA study (i.e., WDL) and low within-person PA variance in one EOD study (i.e., NSDE). For most studies examined, close to half the variance in affect was due to semistable differences across the periods studied between individuals on average. That is, although ILD are often motivated by questions of within-person fluctuation, the widely consistent findings across these datasets indicate that ILD can also provide information about individual differences- and have high reliability with relatively few items. Also, as displayed in Figure S1, although the ratios of between- to within-person variance in PA was somewhat higher in many of the studies that in was in NA, it would not be accurate to say that PA varies less than NA. In absolute terms, only one EOD study (i.e., SAWM) showed greater total variance in NA than in PA. In response to our goal of replication with coordinated analysis, this consistency is notable given the diversity of items, response scales, response options, assessment frequencies, and samples.

Our central question in decomposing affective variance was about the relative contribution of days and moments to the within-person variance that is the focus of much of ILD research on affect. Based on the 5 EMA datasets, we found that about 15% of the total NA variation was at the day level and about 40% was at the moment level. For PA, we found about 13% at the day level and 37% of the total variance at the moment level. These results suggest that the largest proportion of within-person variance arises within days. One possible explanation for this pattern is that the relatively lower day level variance may represent routines and schedules which are fairly consistent when viewed at the day level - that, for example, the overall effect of a given Tuesday and Wednesday may be interchangeable. Over shorter intervals, however, individuals may be shifting environments, social partners, tasks, etc. and these contextual and motivational shifts may help to explain, for example, why an individual's affective experiences at 9 am and 9 pm on the same day may differ (and, in fact, more so than 9 am on one day and 9 am the next day).

We noted earlier that researchers often collect ILD to identify within-person affect variance without distinguishing between the multiple timescales over which these states unfold. We posed the question, "does affect vary more within a day than across days?" in order to assist researchers in both design and analysis choices. For researchers deciding between an EMA and diary approach for affect assessment, we found more within-person NA and PA variance was at the momentary compared to daily level, suggesting that if we relied on a single measurement for each day we may risk that much of the within-person variance we observe is occurring over a faster, unmeasured timescale. For EMA researchers planning analyses and deciding between a 2- (e.g., observation, person) and 3-level (e.g., moment, day, person), we caution, however, that 1315% of the total variance is not negligible. Indeed, about 1/4 of the within-person variance on average was at the day level. If our goal is to understand whether someone is happier or unhappier than usual at a particular time, up to 1/4 of the explanation may depend upon identifying factors that occur over a longer time period (e.g., 24 hours). For example, weekend/weekday and work/non-work effects have been found as predictors of state affect (e.g., Ram et al., 2014; Ryan, Bernstein, & Brown, 2010). Some research questions may involve predictors that occur intermittently across the day (i.e., social interactions, stressors) whereas other predictors may occur only a few times across the study period (i.e., seizure event among individuals with epilepsy, binge drinking in college students) but still have a strong link to affect. Using a 3-level model with EMA data can help to determine how well we have identified predictors of within-person fluctuations in affect.

In the present analyses, we have limited our examination to variance in affective states - how individuals vary from each other on average, how observations from the same individual vary from one day to another, and how observations from the same individual vary from one momentary assessment to another. Affective instability (Ebner-Priemer et al., 2009; Jahng, Wood, & Trull, 2008; Larsen, 1987; Russell & Barrett, 1999; Trull et al., 2008), describing temporal ordering in shifts in affect, is also an important and relevant future direction for replication through coordinated analysis.

Reliability

We replicated prior work showing that end of day and momentary reports show excellent reliability for detecting individual differences ($R_{BP} \geq .95$) using EOD and EMA to assess NA and PA (Cranford et al., 2006). This indicates that, even with varying numbers of items and item content, EOD and EMA studies are well-poised for answering questions of individual differences - such as, do individuals high in Neuroticism report higher NA in daily life than individuals low in Neuroticism?

We also demonstrated adequate-to-high within-person reliability ($R_{WP} \geq .73$). The consistency of the within-person reliabilities across the studies is an important finding. As described in the variance decomposition descriptions, the lowest level of variance estimates (e.g., day-level in diary and hybrid EOD and moment-level in EMA and hybrid momentary) reflect true within-person variation and error. We noted above that a relatively greater proportion of the within-person variance in EMA studies was at the momentary compared to the daily level, thus it is important to examine the reliability of these within-person estimates. The pattern indicates that the within-person variance detected in these datasets was likely due to actual changes in affect states within individuals across time rather than unreliable measurement. Within this pattern, however, it is important to note the relatively lower within-person NA reliability for the diary studies (e.g., WFHS, NSDE). These studies had adequate reliability (R_{WP} = .73-.77) but differed in several ways from the hybrid EOD studies. These lower within-person reliabilities could be due to the reporting of frequency rather than intensity of affective states, the response scale, or the mode of administration (i.e., telephone interview vs. self-initiated mobile survey, number and spacing of assessments). Another possibility is that these studies had many more NA items (1014 compared to 4–5 items) which may have clustered into NA subtypes (see Charles, Mogle, Leger, & Almeida, 2017). That is, pooling across possible subtypes (e.g., anger, sadness) to create a general composite may have resulted in lower reliability in these studies. Within-person reliability estimates for PA in these studies, however, were in line with the estimates from the other EOD data from hybrid studies (i.e., ESCAPE, SAWM) despite the large number of items, use of frequency ratings, and administration mode.

Given the consistency in the estimates for between-person and within-person reliability, one question generated by this work is, do the specific items matter? At least for broad conceptions of NA and PA, it appears the answer is not a lot. Our analyses drew from a total of 23 NA and 26 PA items. We conducted our analyses based on study-specific NA and PA composites. Only happy, tense, sad, and angry (or close variants i.e., extremely happy, tense/anxious, so sad nothing could cheer you up) were shared across 5 or more datasets and none of the studies included all of these items. For the coordinated analysis approach we used, then, examinations of discrete emotions and circumplex models were not appropriate. It is certainly possible that the between- and within-person contributions to variance in, for example, anger-related items may differ from those in depression-related items. By extension, recommendations for numbers of items needed reliably measure a particular subtype of affect (e.g., depression, hostility, anxiety) may differ from the recommendations we provided for general NA and PA. As a starting point for researchers interested in specific subtypes, Cranford and colleagues (2006) calculated reliabilities with three items each for anxious mood, depressed mood, and anger. With the growing popularity of ILD designs, however, it is likely that future work will be able to extend the coordinated analysis approach to datasets with multiple items for specific subcomponents to better address this issue.

Leveraging Coordinated Analyses for Future Design Decisions

The third goal of this project was to aid researchers in designing future studies of affect variance. Bearing in mind differences in sample, items, spacing and total number of assessments, and format, we estimated that the median number of items needed for adequate within-person reliability for NA or PA with a diary or EOD hybrid design is about 3. To reach high (i.e., > .8) reliability, double the items may be needed. For momentary assessments, we found adequate within-person reliability with a median of 2 items. Again, our median estimate indicates that double the number may be necessary to reach high within-person reliability of NA or PA.

These relatively low numbers may be reassuring as researchers grapple with design decisions to balance ILD participant burden with reliability. Based on our results, it appears that a relatively low number of items are likely to achieve high reliability for detecting within-person fluctuations in affect. We note, however, that these items should be carefully chosen if the goal is to create composites and examine constructs like NA and PA. Another important consideration is the balance between psychometrics and content validity. That is, the analyses presented here provide information about the number of items needed for reliable assessments of individual differences and within-person fluctuations. The present work, however, does not shed light on whether the composites (or the specific items) are functionally important in the proposed processes and phenomena (e.g., high vs. low arousal NA items and their signal for predicting health outcomes).

In our secondary analysis of existing data, we were limited to manipulating m (i.e., the number of items) in our calculations and examining the impact on reliability, holding all other contributors in each study constant. We, of course, recognize that many other design factors could impact the other terms in these equations (e.g., the variance due to person, occasion, item, and the interactions of personXitem, occasionXitem, and personXoccasion) and as a result, impact the reliability. For example, as mentioned in the previous section, the two studies with the lowest within-person NA reliability were diary studies (NSDE and WFHS) which asked about the frequency of emotional experiences; all the other studies were phrased as intensity ratings.

These studies had the lowest raw variance in NA and PA (Table S1). Similarly, response format could also influence variance. Only one hybrid study (ESCAPE) used a visual analog scale ranging from 0–100 for EOD and EMA ratings; all other studies used Likert-type ratings with 5–7 response options. As such, there is great need for further psychometric work in ILD measurement, including both secondary analyses and experimental manipulations of design factors (among many other approaches; see for example, Mehl & Conner, 2012; Stone, Shiffman, Atienza, & Nebling, 2007). This psychometric work fits with the National Institutes of Health Common Fund's Science of Behavior Change program's efforts to assist researchers in developing and using verified measures of, for example, stress response indicators such as NA. This growing measures repository is available at https://scienceofbehaviorchange.org/measures/.

Conclusion

These analyses demonstrate that, across studies, about half of total variance can be attributed to between-person differences, with the remaining half due to within-person variance (NA: 45–66%, PA: 25–74%). By utilizing data from EMA studies we were able to demonstrate that substantially more of the within-person variance was attributable to the momentary (within-day) level rather than the daily (between-day) level. Regarding the reliability of assessments, across all studies - despite differences in items, designs, and samples - reliability was adequate to high at all levels of analysis (within-person: .73-.91; between-person: .96–1.00). Moreover, reliable assessments can be readily achieved using relatively few numbers of items, particularly in EMA studies. This work is conducted within the context of growing interest in, and calls for, the use of ILD in clinical settings (e.g., for precision medicine, ambulatory and/or adaptive interventions); yet, ILD approaches lack the psychometric foundation underlying most of our traditional “trait” measures. We hope that this, and related work, will help inform such use of ILD but also serve as a reminder for the need for foundational work (e.g., measurement issues) in these domains.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the National Institutes of Health (NIH) Science of Behavior Change Common Fund Program through an award administered by the National Institutes of Aging (UH2-AG052167). Support for the individual datasets was provided by the sources described below. ESCAPE: This work was supported by the National Center for Advancing Translational Sciences (NCATS), the Leonard and Sylvia Marx Foundation, the Czap Foundation, and the National Institute on Aging (R01 AG039409, R01 AG042595, P01 AG03949, CTSA 1UL1TR001073). NSDE: This work was supported by the National Institute on Aging (P01 AG020166 and R01 AG019239). NTH: This work was supported by the Basic Behavioral and Social Sciences Research Opportunity Network (OppNet) and by the National Heart, Lung, and Blood Institute (R01 HL109340). SAWM: This work was supported by the National Institute on Aging (R01 AG026728). SHADE: This work was supported by the National Heart, Lung, and Blood Institute (R01 HL067990). WDL: This work was supported by the Gallup Organization. WFH: This work was supported by the National Institute of Child Health and Human Development, the National Institute for Occupational Safety and Health, the William T. Grant Foundation, the Alfred P Sloan Foundation, the Administration for Children and Families, the National Institute on Aging, and the National Heart, Lung, and Blood Institute (U01HD051217, U01HD051218, U01HD051256, U01HD051276, U01AG027669, U01OH008788, U01HD059773, R01HL107240).

References

Almeida DM, McGonagle K, & King H (2009). Assessing Daily Stress Processes in Social Surveys by Combining Stressor Exposure and Salivary Cortisol. Biodemography and Social Biology, 55(2), 219–237. 10.1080/19485560903382338 [PubMed: 20183906]

Bisconti TL, Bergeman CS, & Boker SM (2004). Emotional Well-Being in Recently Bereaved Widows: A Dynamical Systems Approach. The Journals of Gerontology: SeriesB, 59(4), P158–P167. 10.1093/geronb/59AP158

Bray JW, Kelly EL, Hammer LB, Almeida DM, Dearing JW, King RB, & Buxton OM (2013). An Integrative, Multilevel, and Transdisciplinary Research Approach to Challenges of Work, Family, and Health. Methods Report (RTIPress), 138. 10.3768/rtipress.2013.mr.0024.1303

Brose A, Scheibe S, & Schmiedek F (2013). Life contexts make a difference: Emotional stability in younger and older adults. Psychology and Aging, 25(1), 148–159. 10.1037/a0030047

Carstensen LL, Turan B, Scheibe S, Ram N, Ersner-Hershfield H, Samanez-Larkin GR, … Nesselroade JR(2011). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. Psychology and Aging, 26(1), 21–33. 10.1037/a0021285 [PubMed: 20973600]

Charles SA, Mogle J, Leger KA, & Almeida DM (2017). Age and the Factor Structure of Emotional Experience in Adulthood. Journals of Gerontology: psychological Sciences, gbx116, 10.1093/geronb/gbx116

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. 10.1177/0146167206287721 [PubMed: 16738025]

Curran PJ, & Hussong AM (2009). Integrative Data Analysis: The simultaneous analysis of multiple data sets. Psychological Methods, 14(2), 81–100. [PubMed: 19485623]

Damaske S, Smyth JM, & Zawadzki MJ (2014). Has work replaced home as a haven? Reexamining Arlie Hochschild's Time Bind proposition with objective stress data. Social Science & Medicine, 115, 130–138. 10.1016/j.socscimed.2014.04.047 [PubMed: 24869785]

de Haan-Rietdijk S, Kuppens P, & Hamaker EL (2016). What's in a Day? A Guide to Decomposing the Variance in Intensive Longitudinal Data. Frontiers in Psychology, 7 10.3389/fpsyg.2016.00891

Diener E, & Emmons RA (1984). The independence of positive and negative affect. Journal of personality and social psychology, 47(5), 1105. [PubMed: 6520704]

Ebner-Priemer UW, Eid M, Kleindienst N, Stabenow S, & Trull TJ (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. Journal of Abnormal Psychology, 118(1), 195–202. 10.1037/a0014868 [PubMed: 19222325]

Eid M, & Diener E (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. Journal of Personality and Social Psychology, 76(4), 662–676. 10.1037/0022-3514.76A662

Gorin AA, & Stone AA (2001). Recall biases and cognitive errors in retrospective selfreports: A call for momentary assessments. Handbook of health psychology, 23, 405–413.

Gross JJ, & Jazaieri H (2014). Emotion, Emotion Regulation, and Psychopathology An Affective Science Perspective. Clinical Psychological Science, 2(4), 387–401. 10.1177/2167702614536164

Hofer SM, & Piccinin AM (2009). Integrative analysis through coordination of measurement and analysis protocol across independent longitudinal studies. Psychological Methods, 14(2), 150–164. [PubMed: 19485626]

Jahng S, Wood PK, & Trull TJ (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. Psychological Methods, 13(4), 354–375. 10.1037/a0014173 [PubMed: 19071999]

Kuppens P, Van Mechelen I, Nezlek JB, Dossche D, & Timmermans T (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. Emotion, 7(2), 262–274. 10.1037/1528-3542.7.2.262 [PubMed: 17516805]

Larsen RJ (1987). The stability of mood variability: A spectral analytic approach to daily mood assessments. Journal of Personality and Social Psychology, 52(6), 1195–1204. 10.1037/0022-3514.52.6.1195

Mehl MR, & Conner TS (2007). Handbook of research methods for studying daily life. New York: Guilford Press.

Mogle J, Muñoz E, Hill NL, Smyth JM, & Sliwinski MJ (2017). Daily Memory Lapses in Adults: Characterization and Influence on Affect. The Journals of Gerontology: Series B. 10.1093/geronb/gbx012

Moskowitz DS, & Zuroff DC (2004). Flux, Pulse, and Spin: Dynamic Additions to the Personality Lexicon. Journal o f Personality and Social Psychology, 86(6), 880–893. 10.1037/0022-3514.86.6.880

Mroczek DK, & Kolarz CM (1998). The effect of age on positive and negative affect: a developmental perspective on happiness. Journal of personality and social psychology, 75(5), 1333 [PubMed: 9866191]

Nezlek JB (2016). A practical guide to understanding reliability in studies of within person variability. Journal of Research in Personality. 10.1016/jjrp.2016.06.020

Pressman SD, & Cohen S (2005). Does positive affect influence health? Psychological Bulletin, 131(6), 925–971. 10.1037/0033-2909.13L6.925 [PubMed: 16351329]

Ram N, Conroy DE, Pincus AL, Lorek A, Rebar A, Roche MJ, … Gerstorf D (2014). Examining the Interplay of Processes Across Multiple Time-Scales: Illustration With the Intraindividual Study of Affect, Health, and Interpersonal Behavior (iSAHIB). Research in Human Development, 11(2), 142–160. 10.1080/15427609.2014.906739 [PubMed: 26989350]

Ram N, & Gerstorf D (2009). Time-structured and net intraindividual variabilit…: Results.Psychology and Aging, 24, 778–791. [PubMed: 20025395]

Ruiz J, Taylor D, Uchino BN, Smith T, Allison M, Ahn C, … Smyth JM (In Press). Evaluating the longitudinal risk of social vigilance on atherosclerosis. Study protocol for the North Texas Heart Study BMJ Open.

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

Russell JA, & Barrett LF (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. Journal of Personality and Social Psychology, 76(5), 805–819. [PubMed: 10353204]

Ryan RM, Bernstein JH, & Brown KW (2010). Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. Journal of Social and Clinical Psychology, 29(1), 95–122.

Scott SB, Graham-Engeland JE, Engeland CG, Smyth JM, Almeida DM, Katz MJ, … Sliwinski MJ (2015). The Effects of Stress on Cognitive Aging, Physiology and Emotion (ESCAPE) Project. BMC Psychiatry, 15, 146 10.1186/s12888-015-0497-7 [PubMed: 26138700]

Segerstrom SC, Boggero IA, Smith GT, & Sephton SE (2014). Variability and reliability of diurnal cortisol in younger and older adults: Implications for design decisions. Psychoneuroendocrinology, 49(Supplement C), 299–309. 10.1016/j.psyneuen.2014.07.022 [PubMed: 25137484]

Smyth JM, Sliwinski MJ, Zawadzki MJ, Scott SB, Conroy D, Lanza ST, …Almeida DM (in press). Everyday stress response targets in the science of behavior change Behavior Research and Therapy.

Smyth JM, Zawadzki MJ, Santuzzi AM, & Filipkowski KB (2014). Examining the effects of perceived social support on momentary mood and symptom reports in asthma and arthritis patients. Psychology & Health, 29(7), 813–831. 10.1080/08870446.2014.889139 [PubMed: 24568534]

Stone AA, Shiffman S, Atienza A, & Nebeling L (2007). The science o f real-time data capture: Self-reports in health research. New York: Oxford University Press.

Suls J, Green P, & Hillis S (1998). Emotional Reactivity to Everyday Problems, Affective Inertia, and Neuroticism. Personality and Social Psychology Bulletin, 24(2), 127–136. 10.1177/0146167298242002

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. Journal of Abnormal Psychology, 117(3), 647–661. 10.1037/a0012532 [PubMed: 18729616]

Watson D (1988). Intraindividual and interindividual analyses of positive and negative affect: Their relation to health complaints, perceived stress, and daily activities. Journal of Personality and Social Psychology, 54(6), 1020–1030. 10.1037/0022-3514.54.6.1020 [PubMed: 3397861]

Watson D, & Clark LA (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.

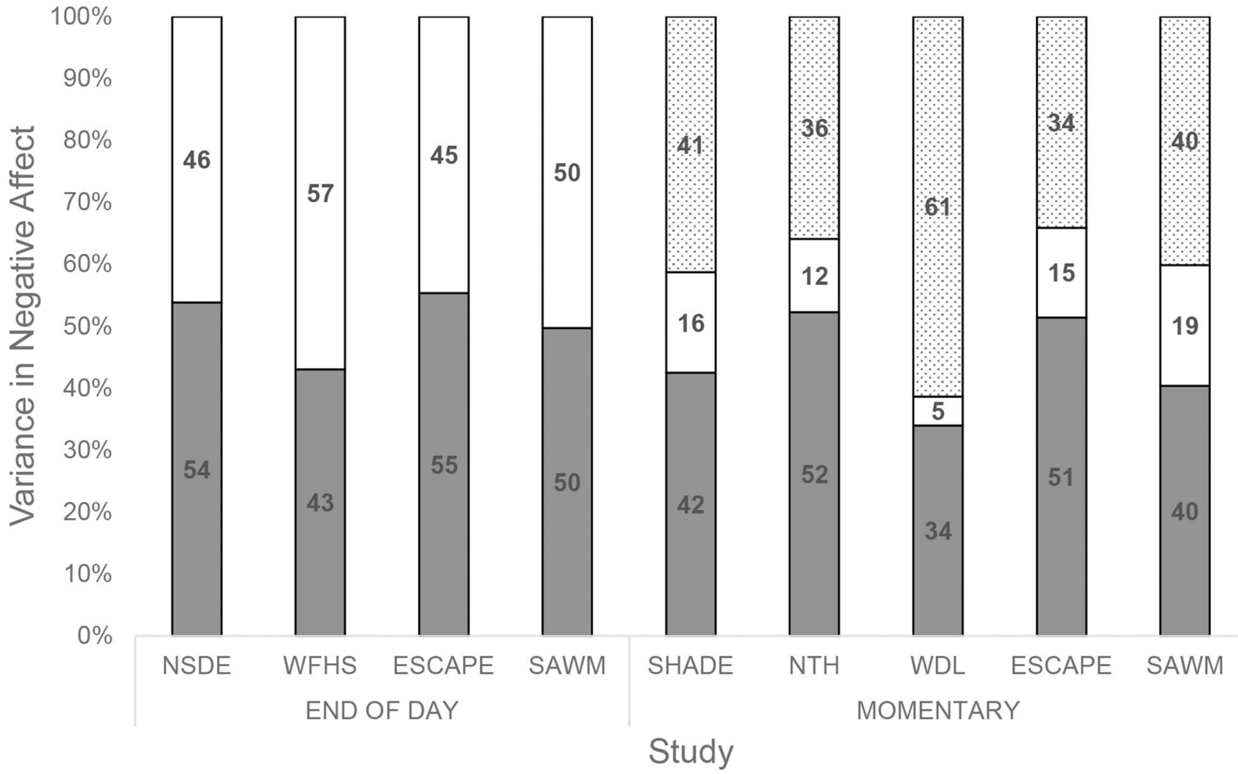

Figure 1.

Percentages of Between (Shaded) and Within (Unshaded) Person Variance in Negative Affect (NA) in End of Day (EOD) and Ecological Momentary Assessment (EMA) Reports. For each dataset, percent between-person variance is displayed in grey shaded portions of the bars and percent within-person variance is displayed in the unshaded portions. In momentary datasets, within-person variance can be decomposed into two sources, variation within-persons across days (the lower portion of the unshaded area) and variation within-persons within days (the upper dotted portion of the unshaded area). ESCAPE and SAWM are hybrid designs containing both EOD and momentary reports, thus the EOD data from the hybrid studies is presented with the diary studies and the momentary data from these hybrid studies is presented with the EMA studies.

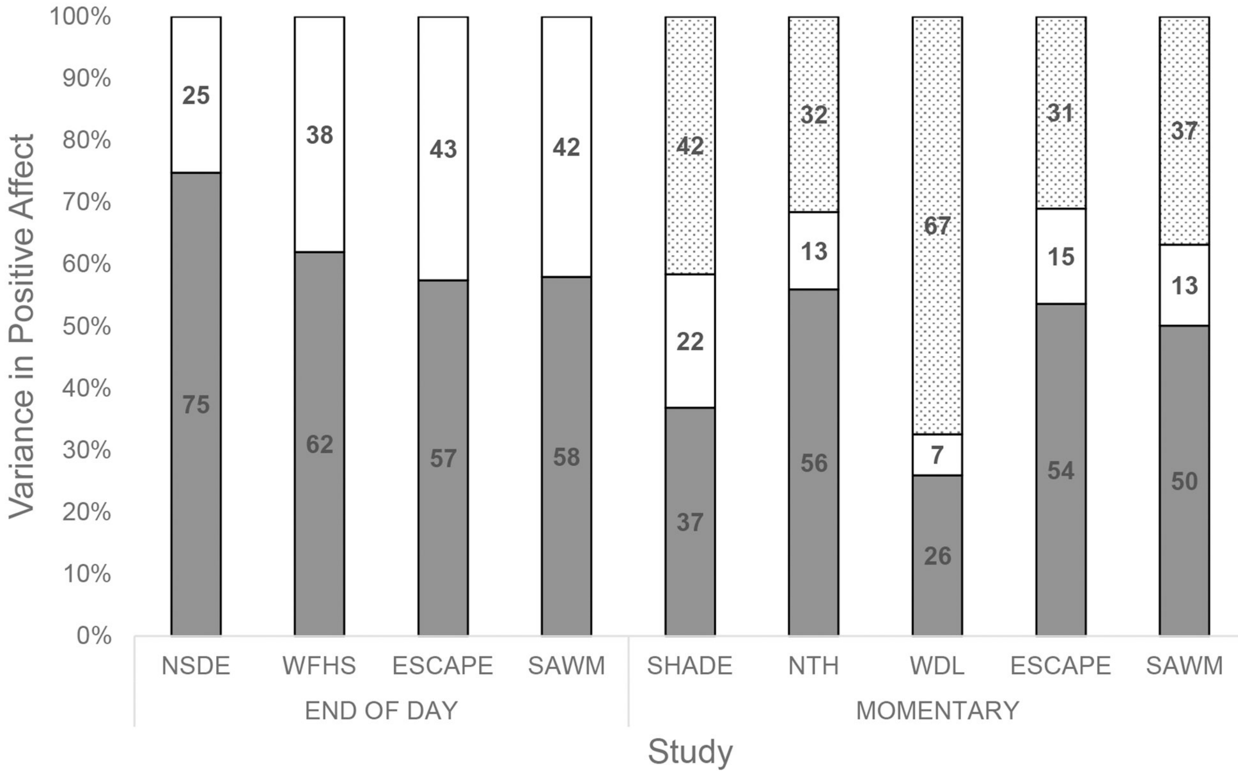

Figure 2.

Percentages of Between- (Shaded) and Within- (Unshaded) Person Variance in Positive Affect (PA) in End of Day (EOD) and Ecological Momentary Assessment (EMA) Reports. For each dataset, percent between-person variance is displayed in grey shaded portions of the bars and percent within-person variance is displayed in the unshaded portions. In momentary datasets, within-person variance can be decomposed into two sources, variation within-persons across days (the lower portion of the unshaded area) and variation within-persons within days (the upper dotted portion of the unshaded area). ESCAPE and SAWM are hybrid designs containing both EOD and momentary reports, thus the EOD data from the hybrid studies is presented with the diary studies and the momentary data from these hybrid studies is presented with the EMA studies.

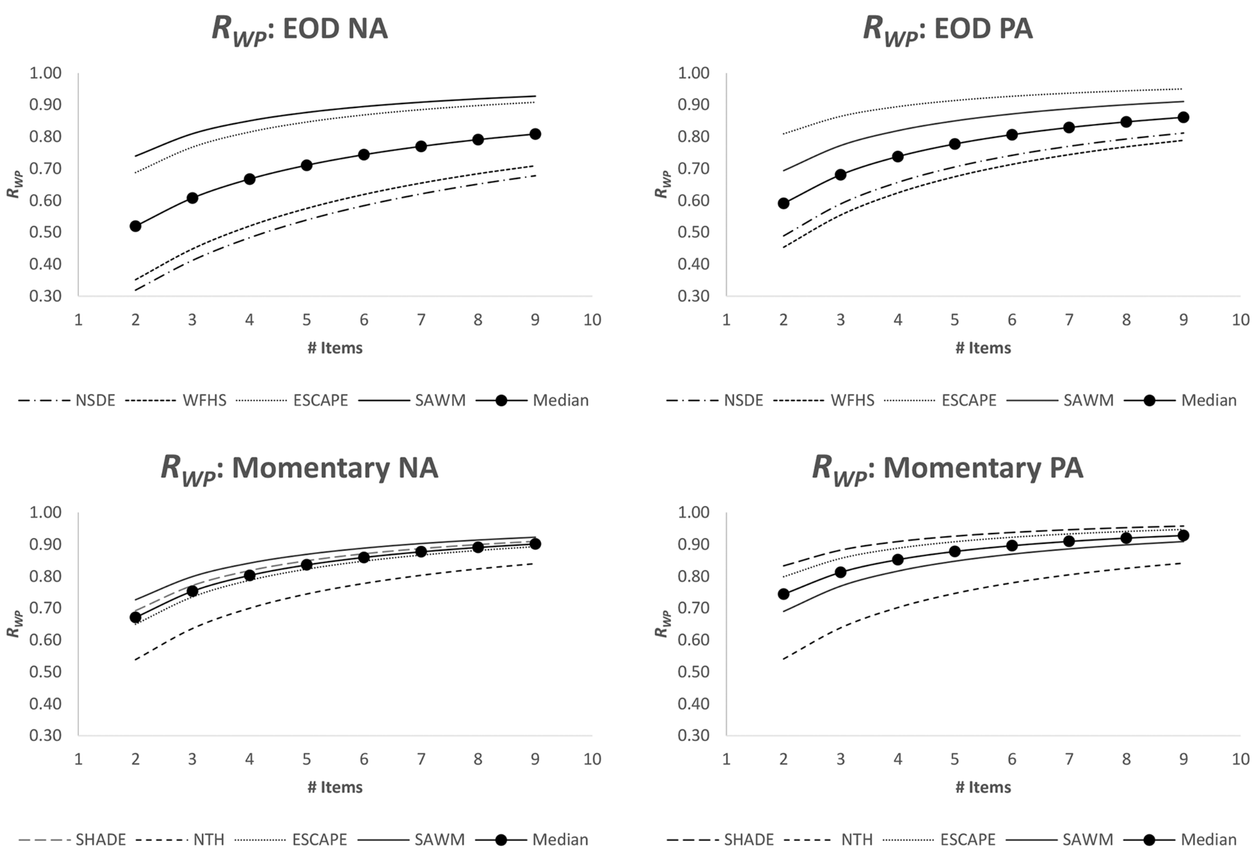

Figure 3.

Plots for estimated reliability of within-person *(R_{WP})* assessments of negative affect (NA) and positive affect (PA) based on coordinated analyses of End of Day (EOD) and Momentary datasets. ESCAPE and SAWM are hybrid designs containing both EOD and momentary reports, thus the EOD data from the hybrid studies is presented with the diary studies and the momentary data from these hybrid studies is presented with the Ecological Momentary Assessment (EMA) studies.

Table 1.

Negative Affect (NA) Items by Study.

	END OF DAY				MOMENTARY				
	NSDE	**WFHS**	**ESCAPE**	**SAWM**	**SHADE**	**NT-Heart**	**WDL**	**ESCAPE**	**SAWM**
afraid	✓	✓							
angry	✓				✓	✓			
angry/ho stile			✓					✓	
ashamed	✓	✓							
depressed					✓	✓			
depressed/blue			✓					✓	
disappointed				✓					✓
distressed		✓							
everything was an effort	✓								
frustrated	✓		✓		✓			✓	
guilty		✓							
hopeless	✓								
hostile		✓				✓			
irritable	✓	✓							
jittery	✓	✓							
lonely	✓								
nervous	✓	✓				✓			
restless or fidgety	✓								
sad				✓		✓	✓		✓
scared		✓							
so sad nothing could cheer you up	✓								
tense				✓		✓			✓
tense/anxious			✓					✓	
tired							✓		
unhappy			✓		✓			✓	
upset	✓	✓		✓					✓
worried					✓				
worthless	✓								

Note. The rows list the negative affect items examined in this coordinated analysis project. The columns list the studies used in this project. A check mark indicates that the item was used in the study. ESCAPE and SAWM are hybrid designs containing both End of Day (EOD) and momentary reports, thus the EOD items from these hybrid studies are presented with the diary studies and the momentary items from these hybrid studies are presented with the Ecological Momentary Assessment studies.

Table 2.

Positive Affect (PA) Items by Study.

	END OF DAY				MOMENTARY				
	NSDE	**WFHS**	**ESCAPE**	**SAWM**	**SHADE**	**NT-Heart**	**WDL**	**ESCAPE**	**SAWM**
active	✓	✓							
alert		✓							
attentive	✓	✓							
calm						✓			
calm and peaceful	✓								
cheerful	✓					✓			
close to others	✓								
confident	✓								
content				✓					✓
determined		✓							
energetic						✓			
enjoyment					✓				
enjoyment/ fun			✓					✓	
enthusiastic	✓	✓		✓					✓
excited		✓		✓					✓
extremely happy	✓								
full of life	✓								
happy			✓	✓	✓	✓	✓	✓	✓
in good spirits	✓								
inspired		✓							
interested		✓					✓		
joyful			✓		✓			✓	
like you belong	✓								
lively						✓			
pleased			✓		✓			✓	
proud	✓	✓							
relaxed						✓			
satisfied	✓								
strong		✓							

Note. The rows list the positive affect items examined in this coordinated analysis project. The columns list the studies used in this project. A check mark indicates that the item was used in the study. ESCAPE and SAWM are hybrid designs containing both EOD and momentary reports, thus the EOD items from these hybrid studies are presented with the diary studies and the momentary items from these hybrid studies are presented with the Ecological Momentary Assessment studies.

Table 3.

Assessment Modes by Study.

	End of Day				Momentary				
	NSDE	**WFHS**	**ESCAPE**	**SAWM**	**SHADE**	**NTH**	**WDL**	**ESCAPE**	**SAWM**
Reporting Frame	Today	Today	Today	Today	Right Now	Right Now	Right Now	Right Now	Right Now
Response Content	Freq.	Freq.	Intens.	Intens.	Intens.	Intens.	Intens.	Intens.	Intens.
Response Scale	Likert	Likert	VAS	Likert	Likert	Likert	Likert	VAS	Likert
Response Options	0=None of the time 1=A little of the time 2=Some of the time 3=Most of the time 4=All of the time	0=None of the time 1=A little of the time 2=Some of the time 3=Most of the time 4=All of the time	0=Not at all 100=Extremely	l=Not at all 4=Moderately 7=Extremely	0=Not at all 6=Very much	l=Not at all 7=Extremely	0=Not at all 6=Very much	0=Not at all 100=Extremely	l=Not at all 4=Moderately 7=Extremely
Source	Kessler et al. (2002), Mroczek & Kolarz (1998)	Watson, Clark, & Tellegan (1988)	Diener & Emmons (1984)	Russel (1980) Watson & Clark (1999)	Diener & Emmons (1984)	Watson, Clark, & Tellegan(1988)	Diener & Emmons (1984)	Diener & Emmons (1984)	Russel (1980) Watson & Clark (1999)

Note. Freq.: Frequency (i.e., how often), Intens.: Intensity (i.e., how much). VAS: Visual Analog Scale. The columns list the studies used in this project. These descriptions apply to both negative and positive affect items described in Table 1. Reporting frame refers to whether participants were instructed to report on their emotions for the day (today) or the moment (right now). Response Content refers to whether participants were instructed to report on the frequency (i.e., how often) or the intensity (i.e., how much) of the emotion. Response Scale refers to whether participants rated their emotions using a Likert-type scale with discrete categories or a continuous slider. Response Options refers to the anchor points and values on the response scale. Source refers to the article or articles from which the emotion items were selected or modified. ESCAPE and SAWM are hybrid designs containing both EOD and momentary reports, thus the EOD items from these hybrid studies are presented with the diary studies and the momentary items from these hybrid studies are presented with the EMA studies.

Table 4.

Between- and Within-person Reliabilities in End of Day Datasets

		End of Day NA				End of Day PA			
	Variance Source	**NSDE**	**WFHS**	**ESCAPE**	**SAWM**	**NSDE**	**WFHS**	**ESCAPE**	**SAWM**
Results	Person	0.04	0.07	286.46	0.75	0.46	0.38	353.28	0.72
	Occasion	0.00	0.01	0.13	0.00	0.00	0.00	−0.55	0.00
	Item	0.02	0.05	15.31	0.06	0.08	0.16	5.53	0.08
	Person*Occasion	0.04	0.07	186.32	0.71	0.15	0.20	221.19	0.47
	Person*Item	0.05	0.06	54.02	0.12	0.19	0.18	17.40	0.24
	Occasion*Item	0.00	0.00	0.37	0.00	0.00	0.00	0.16	0.00
	Error	0.17	0.26	168.99	0.50	0.30	0.48	104.07	0.42
Study Characteristics	k (# occasions)	8	8	14	7	8	8	14	7
	m (# items)	14	10	5	4	14	10	4	4
Reliability Estimates	R_{WP} (within-person reliability)	0.77	0.73	0.85	0.85	0.86	0.81	0.89	0.82
	R_{BP} (between-person reliability)	0.97	0.95	0.99	0.98	0.99	0.98	0.99	0.98

Note. Within-Person (R_{WP}) and Between-Person (R_{BP}) Reliability for Negative Affect (NA) and Positive Affect (PA) across End of Day (EOD) Reports. NSDE and WFHS are traditional daily diary studies; ESCAPE and SAWM results are drawn from the EOD reports from hybrid designs. For clarity, this table presents the results from the in the upper section, followed by the necessary study characteristics needed for reliability calculations, then the reliability estimates.

Table 5.

Between- and Within-person Reliabilities in Momentary Datasets

		Momentary NA				Momentary PA			
	Variance Source	**SHADE**	**NTH**	**ESCAPE**	**SAWM**	**SHADE**	**NTH**	**ESCAPE**	**SAWM**
Results	Person	0.46	0.24	259.99	0.60	0.78	0.73	359.56	0.66
	Occasion	0.01	0.00	0.00	0.00	0.00	0.01	0.24	0.00
	Item	0.17	0.03	11.38	0.06	0.05	0.05	9.85	0.12
	Person*Occasion	0.65	0.19	169.67	0.70	0.93	0.37	247.40	0.61
	Person*Item	0.26	0.11	44.22	0.10	0.09	0.23	20.18	0.24
	Occasion*Item	−0.01	0.00	0.32	0.00	0.00	0.01	0.09	0.00
	Error	0.58	0.33	183.18	0.53	0.37	0.64	124.80	0.55
Study Charaistics	k (# occasions)	35	28	70	35	35	28	70	35
	m (# items)	5	6	5	4	4	6	4	4
Reliability Estimates	R_{WP} (within-person reliability)	0.85	0.78	0.82	0.84	0.91	0.78	0.89	0.82
	R_{BP} (between-person reliability)	0.99	0.99	1.00	0.99	1.00	1.00	1.00	0.99

Note. Within-Person (R_{WP}) and Between-Person (R_{BP}) Reliability for Negative Affect (NA) and Positive Affect (PA) across Momentary Reports. SHADE and NTH are traditional Ecological Momentary Assessment (EMA) studies; ESCAPE and SAWM results are drawn from the momentary reports from hybrid designs. For clarity, this table presents the results in the upper section, followed by the necessary study characteristics needed for reliability calculations, then the reliability estimates. The Work Daily Life (WDL) EMA study is not represented in this table because the study utilized a single item measure for NA and PA, thus neither R_{WP} nor R_{BP} reliability could be calculated.