



Published in final edited form as:

*J Pers Soc Psychol.* 2021 March ; 120(3): 816–835. doi:10.1037/pspp0000289.

## A Direct Comparison of the Day Reconstruction Method (DRM) and the Experience Sampling Method (ESM)

Richard E. Lucas<sup>1</sup>, Carol Wallsworth<sup>1</sup>, Ivana Anusic<sup>2</sup>, M. Brent Donnellan<sup>1</sup>

<sup>1</sup>Department of Psychology, Michigan State University

<sup>2</sup>Supportive.AI, Toronto, Ontario, Canada

### Abstract

The Day Reconstruction Method (DRM) is an approach to measuring well-being that is designed to approximate the rich data that can be obtained from intensive repeated measures designs like those used in the Experience Sampling Method (ESM). Although some preliminary tests of the validity of the DRM have been conducted, these typically focus on agreement between the two methods at very broad levels, rather than focusing on whether the two methods provide similar information about the exact same moments. This paper reports two studies that use ESM and DRM to assess the same moments. Agreement between the two measures varied considerably depending on the focus of the analysis. For aggregate assessments of total time spent in situations and average affect in situations, agreement was high; for between-person differences in time use and experienced affect, agreement varied across situations; and for within-person differences in both situations and affect, agreement was quite low. In addition, we found preliminary evidence that the DRM may be more influenced by expectations regarding the pleasantness of situations as compared to ESM. These results suggest that for many common purposes, the DRM does not provide the same information as ESM.

### Keywords

experience sampling method; day reconstruction method; measurement; affect; subjective well-being

---

Many critical questions in psychology require a detailed understanding of the thoughts, feelings, and behaviors that people experience in specific moments and over time. For example, research examining within-person variability in personality and emotion; research on time use and time management; or research that tracks processes that occur on a very short time scale all require a method for assessing people at several points of time over short intervals. The experience sampling method (ESM; also referred to as ecological momentary assessment) is a well-established and widely used technique for such research (for a review, see Bolger, Davis, & Rafaeli, 2003). Initially administered using paper and pencil measures, advances in survey tools and mobile technology have greatly improved the feasibility of the method. For instance, the ability to signal participants with text messages and to administer

questionnaires by phone greatly enhances the ease of intensive repeated-measures designs. Yet despite these advances, ESM studies remain costly and time-intensive for both researchers and participants. As a result, these methods present methodological challenges that are typically not a concern when simpler methods are used. For instance, the increased respondent burden may lead to higher levels of attrition or to selection biases when ESM methods are used. Thus, in contexts where respondent burden is a concern (such as large-scale survey research that necessarily taps a broad range of content), ESM studies are often not possible.

Partly in response to these methodological concerns, researchers developed the day reconstruction method (DRM; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004). In the DRM, participants are asked to consider the previous day and describe it as a series of episodes, or distinct activities, from the time they woke up until they went to bed. For each episode, participants are asked to report what they were doing, who they were with, and how they were feeling at the time (though the specific questions that are asked can easily be varied). This method was proposed to provide the same degree of data richness as the ESM, while circumventing the need for specialized equipment or software that could be used to administer repeated measurements over time.

Despite the clear benefits of the DRM, this methodology relies on retrospective reports, which can lead to concerns about the accuracy of the information that participants provide. This concern may be especially salient when the construct of interest is a subjective judgment about an internal state (like one's mood or emotions). To understand why, it is important to consider how self-reported evaluations of moods and emotions are made. In their accessibility model of emotional self-report, Robinson and Clore (2002) classify knowledge and reports of emotions into four distinct types. Experiential knowledge, or knowledge of an emotion one is feeling in the moment, is likely to be most accurate (see Kahneman, 1999). Because no recall is required, individuals should be able to introspect about their emotions as they are experiencing them, and reports of these emotional experiences will not be distorted by flawed memories (though, of course, the introspection, itself, may not be perfect).

After a short period of time, the emotional experience passes and respondents must rely on contextual clues—the specific details of the event—to report the associated emotion (Robinson & Clore, 2002). In other words, at this stage, respondents must use episodic memory to provide a report. Because these memories can be inaccurate, recalled levels of emotional experiences may not match what participants would have reported if their experiences had been surveyed in the specific moment. After more time passes, the particulars of an event become even more difficult to remember and individuals may increasingly rely on situation-specific beliefs, or the emotions one would expect to have felt during a particular situation (e.g., funerals elicit sadness). Because these emotions may deviate from expectations in a specific instance, judgments about distant experiences may be particularly prone to error. Finally, when no contextual information is available (as is the case when participants are asked to report on emotional experiences over a long period of time or after a very long delay), individuals might default to identity-related beliefs about

their emotions (Robinson & Clore, 2002). For instance, they may rely on stereotypic beliefs about gender differences in emotional experience.

Under this framework, ESM—at least implementations in which participants are asked to report how they are feeling at a specific moment in time—appears to tap experiential knowledge, as participants are asked to report their emotions as they are experiencing them.<sup>1</sup> The DRM, on the other hand, requires participants to think about emotions from the previous day, which likely requires episodic memory or possibly even situation-specific beliefs that could introduce recall biases and result in lower levels of accuracy. Finally, global reports, which ask participants about their well-being or affect in general, could potentially tap into broader, identity-related beliefs, leading to even weaker convergence within on-line experience. Research supports the conclusion that reports differ at each level of accessibility. For example, when comparing aggregated ESM reports over one week with global reports of affect, Scollon, Howard, Caldwell, and Ito (2009) found correlations of  $r = .52$  for positive affect and  $r = .30$  for negative affect. Hudson, Anusic, Lucas, and Donnellan (2017) found similarly sized correlations when comparing DRM reports of well-being to global reports.

## Previous Research Comparing DRM to ESM

Despite concerns about the accuracy of the memories that are required to make DRM-based judgments, previous work has shown that the DRM produces results that are often similar to those of ESM studies. Specifically, results are similar in regard to typical patterns of affect across time (e.g., diurnal rhythms over the course of a day) as well as reports of typical time usage (Kahneman et al., 2004; Stone et al., 2006). The comparisons in these studies have frequently been indirect, however, with results from the DRM studies compared to the pattern of results reported by participants in separate ESM studies. As Kahneman et al. (2004) state, “Experience sampling is the gold standard to which DRM results must be compared; the DRM is intended to reproduce the information that would be collected by probing experiences in real time” (p. 1777). If the DRM is expected to reproduce the results found via experience sampling, the most direct test of the DRM’s validity would be to compare ESM and DRM reports from the same participants over the same period of time.

To our knowledge, only three studies have conducted such a comparison. Dockray et al. (2010) compared affect ratings derived from the two methods in a sample of employed women on both a work day ( $N = 86$ ) and a leisure day ( $N = 74$ ). Kim, Kikuchi, and Yamamoto (2013) conducted a similar assessment in a small Japanese undergraduate sample (23 men, 2 women). Finally, Bylsma, Taylor-Clift, and Rottenberg (2011) compared a single day of ESM-based affect ratings to those from a DRM assessment in three groups of participants: 35 who were currently experiencing a major depressive episode, 26 in a minor depressive episode, and 38 who had never been depressed. Whereas both Dockray et al. (2010) and Bylsma et al. (2011) found that average affect experienced across the day correlated moderately strongly across the two methods (work day  $r$ s in Dockray et al. = .52

---

<sup>1</sup>ESM instructions that ask participants to describe their experience over longer periods of time may include other forms of knowledge listed by Robinson and Clore (2002).

- .72, leisure day  $r$ s in Dockray et al. = .59 - .79, overall  $r$ s in Bylsma et al. = .62 to .84), Kim et al. found relatively small associations when examining these between-person correlations between average daily affect assessed with ESM and DRM (fatigue  $r = .31$ , depression  $r = .52$ , anxiety  $r = .31$ ). Unfortunately, the relatively small sample sizes mean that these values are estimated with low precision, which makes comparison across studies difficult.

Importantly, several questions about the DRM's ability to reproduce the results of the ESM are as yet unanswered (Diener & Tay, 2013; Tay, Chan, & Diener, 2013). In particular, although Dockray et al. (2010), Bylsma et al. (2011), and Kim et al. (2013) examine cross-method agreement in *average* reports of affect in between-person analyses, these studies did not examine whether the methods agree in other respects. For example, a primary goal of ESM techniques is to trace the timecourse of emotions and situational experiences within a person over the course of a day. However, simple comparisons of average affect cannot determine whether the information that one obtains regarding this time-course is the same or different when using these distinct methods. A more basic question one could ask is whether DRM reports about the specific aspects of a situation (e.g., what a person was doing, who that person was with) provide information that is consistent with what one would obtain from ESM reports. Thus, it is critical to look at agreement within persons over time in addition to looking between persons, as prior work has typically done. Bylsma et al. (2011) did examine within-person associations between events and affective outcomes across the two methods, but they conducted parallel analyses for the ESM and DRM data and did not compare the results directly using statistical techniques.

In addition, given the processes that might be expected to drive participants' responses to questions asked in the context of ESM studies versus DRM studies, it may be the case that there are systematic differences in the responses that participants provide. For instance, in the DRM, participants' recall of their experienced affect in specific situations may be influenced by their expectations regarding how the characteristics of that situation typically influence mood and emotions. As an example, students who retrospectively report on their affect during class may report higher levels of negative affect if they expect to feel more negative in classroom settings. Thus, in addition to examining within- and between-person correlations across the two methods, it is also possible to examine differences in *levels* of reported affect, to determine whether these differ systematically with method.

The present study seeks to address these issues by conducting two studies that directly compare the DRM to the "gold-standard" ESM. In both studies, in addition to examining agreement across the methods in average reports of affect, we also examine agreement in the patterns of affect reported over the course of the day (i.e., within-person), agreement in the situational characteristics reported in each method, and whether reports of affect in each method differ depending on the situational context. In Study 2, we also examine how participants' *expectations* regarding situational effects actually relates to their ratings of emotional experiences.

## Study 1

### Participants

Six hundred sixty-five undergraduate psychology students initially enrolled in the study, which comprised three distinct phases: An initial laboratory session, a day of experience sampling, and a final survey that used the DRM to reconstruct the previous day, during which the ESM reports had been completed. Initial sessions were always scheduled on a weekday, and ESM sessions took place the day after the initial session. This means that all ESM assessments took place from Tuesday through Saturday (with slightly fewer on Saturday than the other days because there were fewer initial sessions scheduled for Friday).

The biggest source of attrition came from those who attended the initial session but then chose not to participate in the ESM and DRM sessions on the next two days. In addition, because the focus of this paper is on the match between ESM and DRM reports, the primary sample is restricted to participants who completed all study phases and who had at least one episode that was reported using both methods. A failure to match occurred when participants did not follow instructions and as a result, failed to provide DRM information on the full day. For example, if a DRM-based episode ended at 11:10, but the participant reported that the next one didn't start until 11:20, the match would fail if the closest ESM report was at 11:15. Four hundred one participants (76% women) with a mean age of 19.49 ( $SD = 1.70$ , range = 18-35) completed all three phases and had at least one matching episode (i.e., the ESM time point had a corresponding DRM episode). However, when examining summary statistics for the ESM and DRM data, we use all available data, including data from 410 participants who completed at least one DRM report and 425 participants who completed at least one ESM report.

Because this study is exploratory in nature, no specific hypotheses were proposed. Instead, our goal was to examine the correspondence between affect and activity ratings assessed using the two methods, and we sought to obtain adequate power to detect relatively small effects. Our goal was to collect as many participants as possible over a two-semester period. With a sample of 401, we have 80% power to detect a correlation of  $r = 0.14$ .

### Procedure and Measures

In the initial session, participants completed a set of questionnaire measures including a single-item measure of life satisfaction, the Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985), and a measure of emotions/feelings/somatic symptoms that had been included in a prior DRM study (Lucas, Freedman, & Carr, 2019). This measure included 9 emotions/feelings/somatic symptoms (happy, frustrated, sad, satisfied, angry, worried, tired, pain, and meaning) that were rated for frequency on a scale from 0, Almost Never, to 6, Almost Always. For some analyses, we report results for individual items, but for most analyses we focus on aggregated positive and negative items. Based on previous analyses of the same set of items in the context of a DRM study (Lucas et al., 2019), we included the items "happy" and "satisfied" as part of the combined positive affect measure and the items "frustrated," "sad", "angry," and "worried" as part of the combined negative affect measure. Because of the ambiguous status of "tired," "pain", and "meaning"

as emotions, we did not include them in the combined measures. Participants in this initial survey were also asked to complete a number of other questionnaire measures about their personality, but these measures were not analyzed for this study (a complete list of questions asked is provided on the associated OSF site: <https://osf.io/dg63b/>).

Participants were asked to provide at least 4 names of people who could provide informant reports of well-being. All informants were emailed and asked to provide confidential reports about the target participant. On average, 2.90 informants per participant provided reports. Informants completed the same well-being measures as participants did, and reports were averaged across informants to create a composite informant for use in all analyses. Additional informant questionnaire measures were administered but not analyzed for this paper; a complete list is again provided at the associated OSF page (<https://osf.io/dg63b/>).

The ESM component of the study was implemented by sending text messages to participants with links to mobile-friendly Qualtrics surveys that participants could complete on their phone. Text messages were sent eight times over the course of a single day. Unique schedules were created for each participant by randomly selecting times within 8 blocks of the day to ensure adequate representation of the entire day. For each moment, participants were asked what they were doing and how they were feeling. Specifically they were asked, “What are you doing?” with 22 response options: Commuting, Shopping/errand, Doing housework, Taking care of your children, Appointment, Working, In class, Studying/doing homework, Preparing food, Eating, Socializing, Intimate relations, Grooming/showering/getting ready, Napping/sleeping, Relaxing, Reading for enjoyment, Entertainment/leisure, Exercising, Watching TV, Computer/internet/email, On the phone, and/or Praying/worshipping/meditating. Response options were not mutually exclusive, and therefore, participants could select more than one option.

Next participants were asked, “Where are you?” with four possible responses: At home, At school, At work, or Somewhere else. Then they were asked, “Who are you interacting with?”, with 10 response options: Nobody, Spouse/significant other, Friends, Roommates, Your clients/customers/students/patients, Co-workers/colleagues/classmates, Boss, Your children, Parents/relatives, and/or Other people not listed (again, more than one response was allowed). Participants were then asked, “Are you currently indoors or outdoors?” (though this question was not analyzed in this paper). Finally, they were asked “How do you feel right now?” and asked to rate the same 9 emotions/feelings/somatic symptoms that were included in the initial survey, this time on a scale from 0, Not at all, to 6, Very much.

The following day, participants were contacted via email to complete a reconstruction of the previous day using the DRM (Kahneman et al., 2004). The DRM was implemented on a web-based platform. Participants were first asked to divide the day into episodes from the time they woke up until they went to bed, providing a start and end time for each, along with a short description to assist with recall. For each episode, participants were then asked questions identical to those used for each ESM time point, with only the tense changed from present to past (e.g., “Where were you?” instead of “Where are you?”).



Studies 1 and 2 were approved as exempt from review by Michigan State University's Institutional Review Board (Protocol X11-703: "Measuring Well-Being").

**Open Science Statement.**—All materials administered in this study, along with the analytic code are available at the Open Science Framework page listed in the author note. Because we did not include a statement about posting data in our informed consent form and because of privacy concerns regarding the identifiability of data due to low-frequency events in the experience sampling and day reconstruction components, data will not be posted publicly. However, Richard Lucas and Brent Donnellan each have archived copies of the data that can be shared with researchers who sign an agreement not to share the data further.

## Analyses

To allow for direct comparisons between the two methods, DRM episodes were first matched with corresponding ESM time points. We considered the two reports to match if the ESM time point fell between the start and end times of a participant's DRM-reported episode. Although this process seems straightforward, there are multiple decisions that must be made when identifying these matches. For instance, decisions must be made about how to treat a single DRM episode that matches multiple ESM reports (e.g., a participant might report being at work from 11:00am to 5:00pm and may have responded to ESM surveys on two occasions during that time). In such situations, we allowed the DRM report to match both ESM reports. To the extent that multiple ESM reports differ within a single DRM episode, this will be reflected in weaker associations across the methods. Identifying these sources of variance are an intended feature of this analysis. Analyses reported below only include participants for whom there was at least one match across methods (2016 matches;  $M = 5$  per participant, range = 1-9).<sup>2</sup>

The results of this study are organized into sections describing multiple questions. First, at the highest level of aggregation (the sample level), we ask whether estimates of total time this sample spent in different situations and the average affect this sample experienced agree across methods. Next, we move down one level of aggregation (to the person level) to focus on the extent to which between-person differences in affect and estimates of time spent in various situations correspond across ESM and DRM methods. This question can be answered by examining correlations between aggregated affect or percent of time spent in each type of situation from all moments of the ESM and all episodes of the DRM. It is important to note that for both sets of analyses, the correspondence across methods would not be expected to be perfect even if both were completely accurate. Even without missing data, the ESM captures eight moments in a day, whereas the DRM could potentially capture affect across all waking hours. However, much research that relies on experiential measures aggregates across moments to obtain a measure that is thought to reflect people's typical experience. If the ESM and DRM are being used by researchers to assess daily average affect and behavior—or even overall trait levels of these characteristics—then the convergence (or lack of convergence) across methods is important to understand.

---

<sup>2</sup>A small number of participants have more than 8 matches because the exact end time of one DRM episode and the exact beginning time of the next DRM episode matched the moment of the ESM report, which resulting in a single ESM report matching more than one DRM report.

Next, we move to the final level of analysis, examining convergence *within people* over time. Specifically, we first look at the extent to which people's within-person patterns of reported affect in the ESM correspond to the within-person patterns found in the DRM. These analyses, which isolate within-person covariation between the two measures across matched moments, determine whether the two methods capture participants' changes in affect throughout the day in similar ways. Again, convergence would not be expected to be perfect, as each DRM episode could potentially capture an event that extends beyond the matching ESM moment that was assessed. However, if these measures will be used by researchers to track the dynamics of affect over the course of the day, it is important to clarify how similar the within-person variance is across methods. Parallel to these analyses, we then ask whether there is agreement within persons over the course of the day about the situations that these participants are in at any given moment. That is, when recalling the previous day's activities in the DRM measure, are participants' reports of situational characteristics accurate? This question can be addressed simply by testing agreement rates across dichotomous items indicating whether a person was or was not in a situation at a particular time. For low frequency activities, where some participants never participate in the activity, within-person agreement cannot be calculated because there is no variance. Even this information is relevant, however, because researchers who are interested in within-person associations will only be able to identify these associations if variability exists. For this reason, when reporting within-person situational agreement, we report both the calculated agreement and the percent of participants who had enough variability to calculate agreement.

Our final question is whether there is an interaction between the method (ESM or DRM) and situational context when predicting affect reports. The goal of these analyses is to determine whether the effect of individual situational factors differs across methods. If, for instance, participants' reports of DRM-based affect is influenced by their beliefs about the impact of situations, then the situational influences may be exaggerated in the DRM as compared to the ESM. The code used for all analyses is included in an Rmarkdown document that is posted on the OSF page for this project: <https://osf.io/dg63b/>.

## Results

We used R (Version 3.6.2; R Core Team, 2019)<sup>3</sup> for all of our analyses.

### Sample-Level Statistics: Cross-Method Convergence in Aggregate Ratings.—

A major goal of this study is to evaluate the extent to which participants provide similar reports across the two methods. Sample-level descriptive information for each of the nine emotions and feelings assessed via ESM and DRM is reported in Table 1.<sup>4</sup> Scores were first aggregated across moments within each participant and then across participants to obtain

<sup>3</sup>We, furthermore, used the R-packages *car* (Version 3.0.5; Fox & Weisberg, 2019), *dplyr* (Version 0.8.3; Wickham et al., 2019), *dummies* (Version 1.5.6; Brown, 2012), *Formula* (Version 1.2.3; Zeileis & Croissant, 2010), *ggplot2* (Version 3.2.1; Wickham, 2016), *Hmisc* (Version 4.3.0; Harrell Jr, Charles Dupont, & others., 2019), *irr* (Version 0.84.1; Gamer, Lemon, & <puspendra.pusp22@gmail.com>, 2019), *lme4* (Version 1.1.21; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.0; Kuznetsova, Brockhoff, & Christensen, 2017), *nlme* (Version 3.1.143; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019), *papaja* (Version 0.1.0.9942; Aust & Barth, 2018), *psych* (Version 1.8.12; Revelle, 2018), *qdap* (Version 2.3.2; Rinker, 2019), *readr* (Version 1.3.1; Wickham, Hester, & Francois, 2018), and *tidyr* (Version 1.0.0; Wickham & Henry, 2019).



sample-level statistics. Means and standard deviations are calculated based on all available data, including data from episodes that did not match across methods and even data from participants who may have completed one type of assessment but not the other. Paired-sample t-tests compare means that include all available data, but these analyses are necessarily restricted to those who had data from at least one moment assessed using each method. Thus, the t-tests reported in the table do not correspond perfectly to the mean differences in the first three columns.

Although we had no reason to expect differences between the two methods, Table 1 shows that when using ESM, participants provided higher ratings of positive and negative affect, all specific positive emotions (i.e., happy, satisfied, meaning), and the negative emotions ‘frustrated,’ ‘angry,’ ‘sad,’ and ‘worried’ than they did when using DRM. No significant differences in responses across methods were found for ‘tired’ or ‘pain’. The general tendency for ratings to be higher when affect was assessed with ESM as compared DRM is consistent with results reported by Bylsma et al. (2011).

Descriptive statistics reported in Table 1 are derived from all available data regardless of whether the ratings came from moments that had been assessed using both methods. This means that differences in results could potentially be due to differences in the way people respond to ESM versus DRM or to systematic differences in the types of moments that are assessed using the two methods. Although we do not report detailed analyses here, we did rerun these analyses focusing only on matched moments, and a similar pattern of results emerged (with significant differences for positive affect, negative affect, ‘happy’, ‘satisfied’, ‘meaning’, ‘sad’, ‘worried’, and ‘tired’). Thus, the differences reported here do not appear to be due to systematic differences in the types of moments that are captured by the two methods.

We next examined sample-level agreement about situational experiences across the two methods. For example, these analyses can determine whether, on average, the estimate of the amount of time that this sample of participants reported being in social situations was similar across methods. In order to answer this question, we first calculated time spent in each type of situation in the form of percentages. For ESM, the number of moments in which a given situation was reported was divided by the total number of moments for that participant. For example, if a participant reported being in a social situation for 3 of their 8 ESM reports, the percentage of time in a social situation would be 37.5%. For the DRM, time-use can be calculated either as a percentage of time spent in a specific situation or as a percentage of episodes in which an activity was reported. For the former, the amount of time spent in each situation was summed and divided by the total amount of time described in the DRM report. For example, if a participant described episodes spanning 10 hours of a day, and indicated that two episodes, each accounting for 1.5 hours, involved social activity, then the percentage of time in a social situation would be 30% (3 hours / 10 hours). To estimate percentage of episodes spent in an activity, the number of episodes in which a person

---

<sup>4</sup>Person-level averages can be computed in one of two ways. Specifically, when calculating an average across an entire day, episodes can be weighted by length or not. Hudson, Lucas, and Donnellan (2019) found that weighting by duration had little effect on the psychometric properties of DRM measures, so here we use simple rather than duration-weighted averages.

reported an activity was divided by the total number of episodes reported. Each of these three indexes reflect subtly different information about how a person spends his or her time, and thus, the absolute levels would not be expected to match perfectly. However, the relative amount of time spent in each activity can be meaningfully compared, and between-person differences should also be comparable.

The average percentage of time participants reported being in each situation for each method is displayed Table 2. The columns will not add up to 100, as the categories were not mutually exclusive. The first thing to note is that despite the fact that absolute levels are not perfectly comparable at a conceptual level, when looking at the day as a whole, estimates of time spent in the various activities are remarkably similar across the two methods. The average absolute discrepancy between the two methods is just 1.70 percentage points when using time spent in DRM episodes as an indicator and 1.70 percentage points when using number of DRM episodes. Discrepancies that did emerge could result from a number of factors including an inability to interrupt the activity to complete an ESM report (which could result in underestimates from the ESM) or to the fact that ESM reports might be scheduled to take place within a more restricted time frame than the DRM, and the types of activities that take place during this restricted time frame may differ from those that take place outside that time frame (which could result in under or overestimates of any activity). Despite these differences, when looking at relative differences across all categories of situation variables, estimates of time spent in each situation from the DRM agreed extremely strongly with those from the ESM. Specifically, the correlation between Columns 1 and 3 from Table 2 was  $r = .97$ , 95% CI [.95, .99],  $t(32) = 24.26$ ,  $p < .001$  and between Columns 2 and 3 was  $r = .97$ , 95% CI [.95, .99],  $t(32) = 23.89$ ,  $p < .001$ . Thus, estimates of the relative amount of time people spend in different situations is quite similar across methods.

In addition to examining estimates of time spent in each situation using the two methods, it is also possible to compare sample-level estimates of experienced affect in each situation. Table 3 reports average positive and negative affect for each of the situation variables assessed in the DRM and ESM. Consistent with Table 1, affect scores from the ESM tend to be higher than those reported using the DRM. However, beyond this discrepancy in absolute level of responses, cross-situational differences in affect are remarkably consistent across methods for both positive and negative affect. The correlation between cross-situational differences in ESM- and DRM-based ratings of affect (i.e., correlations between columns of Table 3) was  $r = .85$ , 95% CI [.73, .93],  $t(32) = 9.32$ ,  $p < .001$  for positive affect and  $r = .76$ , 95% CI [.57, .87],  $t(32) = 6.65$ ,  $p < .001$  for negative affect. Thus, if the goal is to assess, on average, how situations are associated with affective experience, ESM and DRM provide similar information.

**Person-Level Statistics: Between-Person Convergence in Affect and Situations.**—To assess how well ESM and DRM agree regarding *between-person differences* in affect and situation ratings, we first examined correlations between aggregated reports of each emotion. That is, we examined whether DRM-based assessments of a participant's average reported happiness correspond to the same participant's average reported happiness as assessed using ESM. Because we will eventually compare these between-person associations to their within-person equivalents, we restricted these analyses

to moments that were assessed using both ESM and DRM. As shown in the first column of Table 4, these correlations ranged from  $r = 0.76$  (Frustrated) to  $r = 0.90$  (Meaning), indicating a generally high level of between-person agreement across the methods.

We next calculated similar correlations for estimates of the amount of time spent in each situation across the two methods. The first two columns of Table 5 shows these between-person correlations (for time- and episode-based estimates from the DRM). These correlations indicate the extent to which those who report a high percentage of time in a situation using ESM also report a high percentage of time in that situation using the DRM. These correlations vary considerably across situations, ranging from 0.17 (What: Grooming) to a correlation of 0.83 (Where: Work) for time-based estimates and 0.26 (What: Food Prep) to a correlation of 0.76 (Where: Work) for episodes. The average between-person correlation across all situations was  $r = 0.52$  for time and 0.52 for episodes.

**Within-Person Agreement in Affect and Situation Ratings.**—We next examined cross-method agreement at the momentary level. That is, if a participant reports experiencing a high level of happiness in a given moment in the ESM, will that same participant report a similarly high level of experienced happiness when reflecting on that same period of time in the DRM? This question was examined by calculating pooled within-person correlations using the *psych* (Revelle, 2018) package from R.

The second column of Table 4 shows within-person correlations (along with the between-person standard deviations of within-person correlations) between affect measures as assessed by the two methods. The first thing to notice is that in comparison to the between-person associations, agreement within persons over time is much lower. This means that changes in affect as indicated by DRM reports can only be predicted by ESM reports at a relatively modest level. It should be noted, however, that these within-person associations are calculated from at most 8 observations per person. One concern is that over the course of a single day, variability may be limited. Indeed, although we can't use these data to assess whether variability would be greater if we had included a larger number of days, we can compare the within-person variability to the between-person variability that exists in these data. For instance, the within-person standard deviations for positive and negative affect were 0.83 and 0.56 when assessed by ESM and 0.91 and 0.50 when assessed by the DRM. These are somewhat lower than the between-person standard deviations (1.02 and 0.93 as assessed by ESM and 1.16 and 0.94 as assessed by DRM).

As a final test of agreement, it is possible to look within respondents to assess whether specific episodes that have been matched for the time at which they occurred were rated to have the same situational characteristics. To assess this within-person agreement, we calculated kappa coefficients for each person across all matched moments and then averaged to get an overall estimate of within-person agreement. It is important to note that it is not possible to calculate kappa coefficients for respondents who have no within-day variance in a specific situation rating. Thus, someone who consistently indicated in both ESM and DRM that he or she spent no time commuting would not contribute to an overall average agreement rating for the variable "commuting." Similarly, someone who was with friends at every sampled moment would not contribute to agreement statistics even if that person

accurately reported being with friends using both methods of assessment. As a result, the average kappa coefficients that we calculate may underestimate the extent to which the two methods agree. At the same time, if there is no variance within persons, then this variable could not be used to predict other outcomes, which is an important consideration when evaluating the utility of within-person situation ratings in the DRM or ESM. Thus, in addition to calculating kappa, we also report the percentage of respondents for whom this statistic could not be calculated for each of the situation variables.

Table 5 reports the results of these analyses. Specifically, Column 3 reports average within-person kappa coefficients for each situation variable and Column 4 reports the percentage of participants for whom kappa could not be calculated. The first thing to note is that within-person agreement is generally quite low, with an average kappa of 0.34. Second, these values do vary somewhat across situations. The highest agreement is for the location variables (“home”, “school”, “work”, and “other”) and for activity variables of “work” ( $kappa = 0.50$ ) and “class” ( $kappa = 0.72$ ). These latter two variables are those for which respondents may have an explicit schedule, which could facilitate memory.

An important consideration when evaluating these agreement coefficients concerns the fact that many participants do not have enough variance in situation ratings to allow for the calculation of a within-person agreement coefficient. Column 4 shows that this is a concern for many situation variables. Specifically, between 15% and 94% of participants did not report variance in one or both measures, which prevented the calculation of a within-person kappa coefficient. Again, although this makes the interpretation of the averages reported in Table 5 somewhat difficult, it also limits the ability to use these situation ratings to predict within-person changes in other outcomes such as affect.

**Additional Analyses.**—In addition to directly testing agreement across methods, it is also possible to examine the extent to which substantive conclusions differ depending on which measure of experienced affect is used. To address this issue, we examined two additional questions. First, we assessed whether the association between situations and affect variables varied depending which method was used. To conduct these analyses, we restructured the data so that for each matched moment, participants had two lines of data, one reflecting responses to the ESM survey and one reflecting the corresponding variables from the DRM survey. We then tested a multilevel model predicting (in separate models) positive affect and negative affect from one situation variable, the method factor (ESM versus DRM) and their interaction. The intercept and slope for the situation variable were treated as random effects. The critical test is whether the interaction term is significantly different than zero, which would indicate that the association between a situation variable and a respondent’s reported affect varied depending on which method was used.

The results for these interaction tests are presented in Table 6. As can be seen in this table, for the outcome of positive affect, 16 out of 34 interaction terms were significant. For the sake of space, main effects of situations are not reported (these can be found on the associated OSF page: <https://osf.io/dg63b/>), but the general pattern was that for situations that were negatively associated with positive affect, this association was more negative when reported using the DRM than when reported using the ESM. Similarly, for situations that

were positively associated with positive affect, the association was more positive when reported using the DRM than when reported using the ESM. Thus, respondents seemed to accentuate the impact of situations on positive affect when reporting using DRM as opposed to ESM. For negative affect, on the other hand, only 3 out of 34 interaction terms were significant. Thus, although there is some indication that the associations between situation variables and self-reported affective outcomes vary across methods, this evidence is stronger for positive affect than negative affect.

As a second test of whether substantive conclusions differed depending on the method of assessment that was used, we assessed the associations between both self- and informant-report measures of global well-being and the experiential measures derived from aggregated DRM and ESM reports. These correlations are reported in Table 7. For the self-report measures of life satisfaction (both single-item and the Satisfaction With Life Scale) and positive affect, correlations with ESM-based ratings of positive affect were significantly larger than those with DRM-based ratings (Life Satisfaction:  $t = 3.51$ ,  $p < 0.001$ ; SWLS:  $t = 3.87$ ,  $p < 0.001$ ; Global Positive Affect:  $t = 4.32$ ,  $p < 0.001$ ). These differences were not significant, however, for self-reported global negative affect, for experiential measures of negative affect, or for informant-rated criteria. It is also useful to compare these correlations to the correlations between informant reports and self-reported global affect, which are  $r = .37$ , 95% CI [.27, .46],  $t(296) = 6.83$ ,  $p < .001$  and  $r = .35$ , 95% CI [.24, .44],  $t(296) = 6.37$ ,  $p < .001$ , for correlations between global positive affect and informant life satisfaction and SWLS and  $r = -.28$ , 95% CI [-.38, -.17],  $t(296) = -4.92$ ,  $p < .001$  and  $r = -.25$ , 95% CI [-.36, -.14],  $t(296) = -4.50$ ,  $p < .001$  for correlations between global negative affect and informant life satisfaction and SWLS. Although the correlations with global affect are larger than those with ESM or DRM affect, these differences are not significant.

## Discussion

The results of Study 1 suggest that conclusions about the comparability of ESM and DRM measures vary depending on the purposes for which these measures will be used. On the one hand, if the goal is to get sample-level estimates of time spent in specific activities or the average level of affect that people typically report in particular activities, then ESM and DRM provide quite similar information. Similarly, if the goal is to estimate aggregated between-person differences in affective experiences over the course of an entire day, then ESM and DRM correspond reasonably well: Between-person correlations between average affect ratings from the two methods are relatively strong.

Yet despite these forms of agreement, Study 1 also showed that for some commonly used purposes, ESM and DRM do not provide equivalent information, at least when assessed over the course of a single day. Specifically, estimates of between-person differences in time use varied quite dramatically across different situational variables. For instance, although correlations between person-level estimates of time spent in specific locations reached as high as 0.83, the correlation between estimates of a relatively frequent activity—time spent eating (which makes up approximately 14% of participants' time) across the two methods was just 0.21.

Similarly, when moving to the within-person level, where changes in situations and affect over the course of the day are the focus, cross-method agreement was especially weak. This lack of agreement is important because one of the primary goals behind the development of experiential measures is that such measures could be used to determine how changes in experiences over an extended period of time could be linked with changes in experienced well-being. Our results suggest that some caution is warranted when using experiential measures for these purposes (though our design cannot determine whether both ESM and DRM methods suffer from problematic psychometric characteristics, whether only one is problematic, or whether both are reliable and valid but measure different aspects of experience). The weak within-person agreement across these two methods does suggest, however, that they are not interchangeable and that conclusions about substantive questions could differ depending on the method that is used.

One caveat that must be mentioned is that limits in the amount of variance that exists in situational or affective measures over the course of a single day of experience may contribute to these reduced correlations. Indeed, our analyses show that especially for many situational variables, a large proportion of respondents simply did not have any variance with which to calculate agreement statistics. This might suggest that extending the length of experiential measurement sessions beyond a day could improve agreement. However, DRM measures were specifically developed to be included in large-scale, national surveys (Kahneman et al., 2004), and many implementations do cover a single day's worth of experiences (e.g., Stone et al., 2006; Bylsma et al., 2011; Daly, Delaney, Doran, Harmon, & MacLachlan, 2010; Lee, Tse, & Lee, 2017; Oerlemans & Bakker, 2014b; Srivastava, Angelo, & Vallereux, 2008). Researchers who use these methods to examine within-person changes over the course of the day should consider whether the data that result from these experiential measures provides information of sufficient quality for the intended questions.

## Study 2

The goal of Study 2 was to replicate the results of Study 1 using a slightly different method for collecting ESM and DRM data. Specifically, respondents were provided with a single app that administered all three components of the study: The initial survey, up to eight experience-sampling surveys, and the final day reconstruction. The consistency of mode of administration was intended to maximize ease of use, which we hoped would reduce data loss. In addition, the DRM survey enforced logical start and end times for episodes, which should improve our ability to match ESM moments with DRM episodes. Finally, we added a questionnaire to assess participants' global expectations about the impact that specific situational factors had on their mood, with the aim of assessing whether these expectations moderate the effect of situations on affect, especially in interaction with the specific method of assessment. If DRM-based surveys rely more on systematic beliefs about situations and less on episodic memories than do responses to in-the-moment ESM-based surveys (Robinson & Clore, 2002), then it is possible that these expectation measures will interact with the method factor when predicting self-reported affect.



## Participants

A sample of 589 participants completed the initial session, which included a survey with measures of personality and global well-being. After completing the initial survey, participants were then given instructions about the timing and completion of the ESM and DRM surveys. The requirements of the study were to complete one day of ESM, followed by a DRM report of activities over the course of the same day. However, the app also provided feedback to participants about their moods and activities, and participants were allowed to keep using the app after their day of participation. In addition, because of technical problems where participants were prevented from logging in to the app, some participants completed multiple partial sessions before getting a complete assessment. For these reasons, we selected the first DRM session that participants completed, along with any ESM reports that fell during the period covered by this DRM session. As in Study 1, initial sessions took place on weekdays and most initial ESM sessions were completed the following day, which means that most ESM sessions took place on Tuesday through Saturday (though because of scheduling issues, the most frequent days sampled were Tuesday, Wednesday, and Thursday). In Study 2, however, the first day of ESM that overlapped with a DRM occasionally occurred at least one day after the initial session, which means that a small number of ESM sessions were completed on Sunday or Monday.

Of the initial participants, 461 reported on at least one DRM episode during their first DRM session and 419 completed at least one ESM report during this time. For all participants who completed an ESM report during the DRM period, at least one ESM report could be matched with a DRM episode, with an average of 6.49 matches per participant (range = 1 to 8). Thus, the use of the app was successful increasing the number of matches across ESM and DRM reports as compared to Study 1. Those participants who completed all three portions of the study make up the primary sample. Of those reporting their gender, 75% were women; the average age was 20.48.

## Procedure and Measures

The procedures and measures were very similar to those from Study 1. First, participants who indicated that they had an iPhone or Android device were invited to an initial session, where they were asked to download the app on which data collection would take place. The structure of the study was described, and a brief video explaining how to use the app was played. Next, participants completed an initial survey that included measures of personality, general well-being (including all measures from Study 1) and a few measures that were included for other purposes (for a full list of items see materials on our OSF page: <https://osf.io/dg63b/>). In addition, for each situational question included in the ESM and DRM surveys (including questions about what participants were doing, who they were with, and where they were), participants were asked how much they thought these situations typically affected their mood on a seven-point scale that ranged from “affects me very negatively” to “does not affect me” to “affects me very positively.” Unlike Study 1, no informant reports were obtained in this study.

Next, experimenters explained that over the course of the following day, the app would signal participants a total of eight times. When signaled, participants were asked to complete

a brief survey with the same questions that were used in Study 1. Finally, on the day following the ESM surveys, participants were again asked to complete a DRM survey using the app. The same questions from Study 1 were included on the DRM survey.

**Open Science Statement.**—All materials administered in this study, along with the analytic code are available at the Open Science Framework page listed in the author note. Because we did not include a statement about posting data in our informed consent form and because of privacy concerns regarding the identifiability of data due to low-frequency events in the experience sampling and day reconstruction components, data will not be posted publicly. However, Richard Lucas and Brent Donnellan each have archived copies of the data that can be shared with researchers who sign an agreement not to share the data further.

## Analyses

The analyses for Study 2 mirrored those from Study 1 with one exception. Because we included questions about participants' expectations for situations, we also calculated the correlation, across all 31 situations that were presented to participants in the expectation questionnaire, between expectations and average positive and negative affect in the ESM and DRM measures. In addition, we examined the extent to which expectations regarding each situation moderated the within-person changes in aggregated positive and negative affect that respondents reported as they moved between situations. Because we were most interested in assessing people's relative evaluations of situations, we centered each person's situational expectation variables around their overall mean for all situations when using them as a moderator in the multilevel models that tested these within-person effects. These analyses were conducted for each method separately.<sup>5</sup> The syntax used for these models is included in the rmarkdown document on the corresponding OSF site (<https://osf.io/dg63b/>).

## Results

As in Study 1, results are presented in sections corresponding to the questions of interest. We first focus on whether conclusions about sample-level estimates of time use and average affect in situations differ depending on method. We then examine whether between-person differences in affect and time use are similar across methods. Next, we test correspondence between within-person changes in affect and situational experiences across the two methods. As in Study 1, these tests are followed by an examination of method by situation interactions when predicting within-person changes in affect. As an addition to the analyses conducted in Study 1, we examine the links between situation expectations and affect. Finally, as in Study 1, we compare aggregated ESM and DRM reports to global reports of SWB from the initial survey.

**Sample-Level Statistics: Cross-Method Convergence in Aggregate Ratings.**—Table 8 shows descriptive statistics for each of the nine emotions assessed and the combined positive and negative affect scores for ESM and DRM separately. In addition, this table

---

<sup>5</sup>We also planned to conduct a combined analysis that tested higher-order interactions between method and expectation; but as described below, few interactions with expectations emerged in the method-specific analyses, which meant that these more complex follow-up tests were not needed. In addition, preliminary analyses suggest that power to detect such higher-order effects would be quite limited.

shows effect sizes for the absolute differences in ratings across methods and statistical tests of these differences. As in Study 1, some differences did emerge, with affect scores sometimes (though not always) higher in the ESM ratings as compared to DRM. However, these differences are generally smaller in the current study (average  $d = 0.08$ ) as compared to Study 1 (average  $d = 0.14$ ). One possible explanation for this cross-study discrepancy is that the differences found in Study 1 were not the result of general differences in responses to day-reconstruction methodology as compared to experience sampling, but to differences in mode of administration. In Study 2, the use of a single app allowed for greater consistency in mode of administration, which could have reduced differences in reported affect across the two methods.

Table 9 shows percent of time (for DRM) and percent of episodes (for DRM and ESM) for which each situation variable was endorsed. As was true in Study 1, overall agreement across methods in these aggregated time-use estimates is quite high. Correlations in time-use estimates across methods were again very strong:  $r = .86$ , 95% CI [.74, .93],  $t(32) = 9.51$ ,  $p < .001$  for duration of DRM episodes and  $r = .94$ , 95% CI [.89, .97],  $t(32) = 16.25$ ,  $p < .001$  for number of DRM episodes. In fact, even across samples, agreement is high, with strong correlations of estimates of time spent in each situation for DRM duration ( $r = .90$ , 95% CI [.81, .95],  $t(32) = 11.59$ ,  $p < .001$ ), percent of DRM episodes ( $r = .96$ , 95% CI [.92, .98],  $t(32) = 19.37$ ,  $p < .001$ ), and percent of ESM episodes ( $r = .98$ , 95% CI [.97, .99],  $t(32) = 30.70$ ,  $p < .001$ ). Absolute estimates are again, often very similar across methods, though as noted in the discussion to Study 1, comparisons of absolute estimates are difficult to interpret because technically, they are estimating different amounts. Despite this, absolute differences were just 3.64 percentage points for duration and 2.70 for episodes. Study 2 confirms that information about sample-level statistics about average time use are quite consistent across the two methods.

We also estimated the average affect experienced in different situations to determine which situations are most and least pleasurable. Table 10 shows average situation-specific positive and negative affect scores separately for DRM- and ESM-based ratings. Average cross-situational differences in affect are quite consistent across methods for positive affect ( $r = .93$ , 95% CI [.86, .96],  $t(32) = 13.98$ ,  $p < .001$ ), but less so for negative affect ( $r = .57$ , 95% CI [.28, .76],  $t(32) = 3.89$ ,  $p < .001$ ). In Study 1, these correlations were quite similar (and quite high) for both positive and negative affect, so the moderate correlation for negative affect is a discrepancy from Study 1. A closer look at Table 10 shows that the weaker correlation is due partly to a relatively large discrepancy in negative affect scores across the two methods for the activity of reading. However, even without this data point, the correlation is still weaker than in Study 1 ( $r = .70$ , 95% CI [.47, .84],  $t(31) = 5.48$ ,  $p < .001$ ). The weaker correlation may be due to the fact that there was much less cross-situational variability in negative affect ratings compared to positive affect:  $SD_{DRM} = 0.46$  and  $SD_{ESM} = 0.35$  for positive affect versus  $SD_{DRM} = 0.22$  and  $SD_{ESM} = 0.20$  for negative affect.

#### **Person-Level Statistics: Between-Person Convergence in Affect and Situations.**

Column 1 of Table 11 shows the between-person correlations for the mood scores (including the aggregated positive and negative affect scale scores). These values are consistent with correlations from Study 1 (average  $r$  from Study 1 = 0.83, average  $r$  from

Study 2 = 0.79). These values show that estimates of affect reported over the course of the day for different individuals are quite similar across methods.

The first two columns of Table 12 show the between-person correlations between ESM and DRM measures of time spent in each situation (for both duration of DRM episodes and number of DRM episodes). Consistent with Study 1, there is considerable variability in rates of between-person agreement. In this case, correlations range from a low of .10 for time spent worshipping to a high of .86 for time spent at work, with an average correlation of 0.51 for percent of time and 0.50 for percent of episodes. The situations for which agreement is high versus low was consistent across studies: The correlation between the between-person correlations in Study 1 and those in Study 2 is  $r = .84$ , 95% CI [.70, .92],  $t(32) = 8.65$ ,  $p < .001$  for time and  $r = .79$ , 95% CI [.62, .89],  $t(32) = 7.29$ ,  $p < .001$ . Thus, for some situations, between-person differences in time use are estimated reasonably well by the DRM and ESM, whereas for others, agreement coefficients are quite low.

**Within-Person Affect and Situation Ratings.**—Within-person correlations between the two methods of assessing affect (ESM and DRM) are shown in Column 2 of Table 11. As was true in Study 1, these correlations are generally small to medium in size (average  $r$  from Study 1 = 0.30, average  $r$  from Study 2 = 0.26). These relatively weak correlations demonstrate that moment-to-moment changes in affect over the course of a day differ across methods. Again, however, it is important to caution that these within-person associations are derived from an average of just over 6 (and, at most, 8) moments per person. Thus, they could be impacted by low variability over the course of a single day.

Table 12 shows the average within-person kappa coefficients separately for each situation variable. These coefficients reflect average within-person, cross-method agreement about whether a person was in a specific situation at a particular moment. As in Study 1, these agreement coefficients are generally quite low, ranging from .04 for worship to .64 for being in class. The average kappa coefficient across all situation variables was just 0.29. Again, it is important to reiterate that these values need to be interpreted in the context of the within-person variability that exists in these situation ratings. As Column 4 shows, for many situations, many participants do not have enough within-person variability to estimate kappa coefficients. As was the case in Study 1, situation variables that reflect locations or highly scheduled activities (like being at work or in class) exhibited the highest agreement.

**Method Interactions.**—As in Study 1, we also tested whether the association between specific situation variables and reports of positive and negative affect varied depending on the method that was used. These analyses tested separate multilevel models, predicting positive and negative affect from one situation variable, a method factor, and their interaction. The critical result is whether the interaction was significant, and the estimates for these interactions are presented in Table 13 (main effects are again reported on our OSF page: <https://osf.io/dg63b/>). As was true in Study 1, more significant interaction terms were found for ratings of positive affect than negative affect. Importantly, a similar pattern of results emerged across the two studies. Of the 16 significant interaction effects found in Study 1, 11 (those for class, homework, socializing, sleeping, relaxing, entertaining, TV, colleagues, home, school, and other) were also significant in Study 2. In both studies,

participants reported more negative associations with less pleasant situations and more positive associations with more pleasant situations when using the DRM as compared to ESM.

**Expectations.**—A novel feature of Study 2 was that participants rated their expectations regarding the effects of specific situations on affect. To test whether these expectations match with people's experience, we correlated average expectations for each situation with actual experienced positive and negative affect from the DRM and ESM. These correlations were quite strong ( $r_s = 0.74, -0.76, 0.66, -0.63$ , for DRM positive affect, DRM negative affect, ESM positive affect, and ESM negative affect, respectively), suggesting that people have a reasonable understanding of how their affect changes across situations, or perhaps that these expectations even guide experiential reports. The fact that the correlations between expectations and DRM measures are slightly higher than those with ESM measures is consistent with our prediction that expectations may influence recall even in experiential measures like the DRM. It is important to note, however, that because these correlations reflect covariation across just 31 situation variables, the test of differences across methods is under-powered. Indeed, the difference between correlations with the DRM and ESM was not significant for positive affect ( $t = 1.75, p = 0.092$ ) or negative affect ( $t = -1.18, p = 0.249$ ). It is important to note that power to detect these specific sample-level associations cannot be increased by adding participants, as the sampling unit is situations. In addition, adding situations may not help, because it may be difficult to incorporate additional situations that are experienced frequently by participants (and thus add meaningful variance). Therefore, robustness of these differences will need to be tested in additional replication studies.

We also examined whether within person changes in affect were moderated by individual-level expectations regarding these situations. Specifically we tested separate multilevel models for each situation predictor, both positive and negative affect outcomes, and each method of assessment, to see whether person-centered expectation variables moderated the effect of situation on affect. This large set of analyses resulted in a very small set of significant interactions, so we do not report them here (they are available on our OSF page: <https://osf.io/dg63b/>). Instead, we will simply note that out of 31 interactions tested for positive affect in the DRM, only four were significant: being in class, sleeping, being with boss, and being at work. Out of 31 interactions tested for negative affect in the DRM, only 1 was significant: being in class. Of the 31 interactions for positive affect in the ESM, four were significant: being in class, doing food preparation, being with boss, and being home. And finally, of the 31 interactions for negative affect in the ESM, four were significant: doing housework, doing food preparation, being with friends, and being with roommates. Thus, there is little evidence that individual differences in expectations moderate within-person changes in affect as people move to different situations; however, all of the issues we raised about low variability in both situations and affect within the course of a day likely make power to detect these effects quite low. Unlike the sample-level analyses described in the previous paragraph, for these analyses, future research could increase power by adding participants or additional days of assessment.

**Additional Analyses.**—As a final analysis, we again correlated between-person differences in affect from the DRM and ESM with global ratings of SWB from the initial questionnaire. Table 14 shows these results. Consistent with Study 1, correlations were slightly higher for ESM measures than DRM measures, with these differences being significant for correlations between experiential measures of positive affect and the single item measure of Life Satisfaction ( $t = 3.35, p < 0.001$ ) and SWLS ( $t = 3.12, p = 0.002$ ). Unlike in Study 1, no other correlations were significantly different across the methods.

## Discussion

Many results from Study 1 replicated well in Study 2. First, as in Study 1, at the highest level of aggregation, agreement across the two methods was quite high: Estimates of the amount of time people typically spend in each situation and the affect that people typically experience in these situations were very similar across methods (though correspondence in average negative affect experienced in specific situations was somewhat lower in Study 2 than in Study 1). This result is consistent with the initial validation evidence that Kahneman et al. (2004) provided when first proposing the DRM. In addition, the pattern of between-person correlations, both those involving self-reported affect and those involving self-reported situational characteristics, was consistent across studies. Between-person correlations for affect were quite high in both studies, and between-person correlations for time spent in different situations varied in similar ways across studies. Again, these variations suggest that at the person-level, correspondence across methods varies considerably across situations, which means that information about specific people's experiences can diverge depending on the method that is used. Third, in both studies, within-person agreement—both in terms of situational and affective experiences—was quite low. One implication of this finding may be that when within-person processes are of interest, ESM and DRM may give different results. Finally, the interactions between method and situation showed that the impact of many situations on affect ratings was stronger for DRM than for ESM, though this effect emerged primarily for ratings of positive rather than negative affect, and it did not occur for all situation variables. Many of the interactions that did emerge, however, were consistent across studies.

Study 2 also went beyond Study 1 by including direct measures of participants' expectations regarding the affect they would experience in each situation. At the highest level of aggregation (i.e., aggregated across individuals), these single-item expectation variables correlated quite strongly with the actual affect that people experienced, suggesting that, on average, people have some understanding of how situations are associated with affective experience. Indeed, the very strong correlations suggest that actually assessing experiences may not be needed if one's goal is simply to rank how pleasurable different situations are. This information is reflected quite well in respondents' self-reported expectations regarding specific situations. In addition, although these analyses were underpowered and thus not significant, there was suggestive evidence that expectation variables correlate more strongly with DRM reports than with ESM reports. In combination with the finding that the associations between affect and the various situation variables were often significantly stronger for the DRM than the ESM, this suggests that participants may be relying on their



understanding of the typical effect of situations more when using the DRM than the ESM (Robinson & Clore, 2002).

We were also able to examine the extent to which situational expectations moderated within-person changes in affect as people went from one situation to another, but very few of these effects were significant, and those that were were not consistent across methods or affect variables. Thus, although *average* expectations for situations appear to be correlated with *average* levels of affect (especially when assessed using the DRM), individual-level expectations do not predict person-specific changes in affect from situation to situation. Again, however, caution is needed when interpreting these results, due to the fact that our earlier analyses suggest that within-person changes in affect and situations are not estimated reliably with these techniques.

## General Discussion

Understanding the thoughts, feelings, and behaviors that people experience in specific moments and over short periods of time is an important goal for many research programs. More and more methods for assessing these momentary experiences have been developed, yet not all have been subjected to rigorous tests of reliability and validity. The DRM represents a potentially important methodological advance over other measures of momentary experience because it provides within-person data using techniques that can be flexibly incorporated into a variety of research settings that are not amenable to more traditional, time-intensive repeated measures designs. Indeed, precisely because of the advantages that the DRM has in terms of respondent burden, a number of large-scale surveys, including the American Time-Use Survey, the German Socio-Economic Panel Study, and the Panel Study for Income Dynamics have incorporated some form of the DRM into their studies. The goal of this paper was to investigate previously ignored issues regarding the validity of the measures that can be derived from this promising approach.

The primary conclusion from our two studies is that the validity of measures derived from the DRM depends critically on the purpose for which the measures will be used. If researchers are simply interested in obtaining broad, sample-level estimates of time spent in specific situations and the typical levels of affect experienced in those situations, then the information that can be obtained from the DRM is virtually identical to that obtained from ESM. It should be noted, however, that at least for ratings of situation-specific affect, the information derived from the DRM is also quite similar to what respondents provide when they are simply asked how they *expect* to feel in a given situation. Although the DRM may be less burdensome than ESM, it is certainly more time consuming to complete than are simple, global survey responses; and this finding leads to questions about the added benefit of the DRM (or even ESM) if one's goal is simply to provide situation-level aggregates of affect and time-use.

Second, if the research question focuses on between-person differences in affect, then again, the DRM and ESM have high levels of agreement and seem to provide similar information. Once responses are aggregated over an entire day, DRM-based estimates of positive and negative affect correlate strongly with ESM-based estimates. In addition, correlations

between these aggregated affect ratings and global self-reports of affect were only moderate, suggesting that the information obtained from experiential measures, even at the aggregate level, is not identical to what would be obtained from more traditional global measures.

Interestingly, the benefits of aggregation do not always carry over to estimates of between-person variation in time spent in different situations, at least not for all situations. Across both studies, agreement for aggregated, between-person levels of time spent in various situations was moderate on average and varied considerably across situations. Even ratings of some relatively frequent behaviors, like eating, showed low levels of agreement. Thus, some caution is warranted when using person-level time-use estimates from the ESM and DRM. Importantly, our studies cannot reveal whether, to the extent that these two measures should be expected to provide the same answer, the DRM lacks validity, the ESM lacks validity, or both are problematic.

Finally, and perhaps most importantly, both studies consistently showed that within-person agreement—both for affect and situation ratings—was low. In other words, when assessing how people change from moment to moment, the ESM and DRM provide different information about the nature of the changes that occur. Of course, well-known effects of aggregation on the reliability of assessment would lead to the expectation that agreement would be at least somewhat lower for within-person estimates than for aggregated between-person estimates (e.g., Epstein, 1979). Given that a primary goal of intensive repeated-methods methodology like the DRM and ESM is to allow for the study of within-person processes, however, the fact that agreement about the within-person changes that occur is so low should be concerning for people who use these measures.

### Limitations

It is important to acknowledge one major limitation of our studies: In both studies only a single day of experience was sampled. This is important because the limited number of sampled occasions means that variability in some predictors and outcomes might be low. For instance, many participants had no variability whatsoever in some situation variables, meaning that cross-method agreement coefficients could not be estimated. Thus, it is possible that agreement coefficients from our studies may underestimate the agreement that would have been obtained had more days of assessment been used.

However, it is also important to note that our methods mirrored those that are typically used when the DRM is implemented. Most studies that have used the DRM, including both large scale studies like the German Socio-Economic Panel Study, the American Time Use Survey, and the Panel Study for Income Dynamics, and much smaller, investigator-driven studies (e.g., Daly, Baumeister, Delaney, & MacLachlan, 2014; Bylsma et al., 2011; Lee et al., 2017; Oerlemans & Bakker, 2014a; Srivastava et al., 2008; Stone et al., 2006) typically assess one or two days of experience per wave. Indeed, the primary advantage of the DRM approach is that respondents can complete the DRM in a single survey session, using the same mode of administration and at the same time that they complete various other types of survey questions that are typically asked. This means that repeated assessments—which would be required for repeated DRMs—are often not feasible. Thus, although our results

may not generalize to more optimal administration procedures, they should generalize to many existing implementations, including those from widely analyzed studies.

Furthermore, it is not clear how many days of assessment would be required to obtain substantial within-person variability in these measures. To address this issue, we turned to an additional dataset that included three days of DRM assessments (Hudson et al., 2019). Although ESM measures were not available in this study, so we could not examine the effect of using more days on cross-method agreement, it is possible to see how within-person variability changes with additional days of assessment. In this additional dataset, increasing the number of days from one to three only lead to about a 5% increase in within-person standard deviations for the affect variables, and approximately 13% larger within-person standard deviations for situational variables. Thus, it is possible that the low within-person variance reflects the reality of people's experience and is not a result of relying on a single day of assessment. If so, then assessing more days of experiential measures may not lead to greater cross-method correspondence.

An additional limitation of our study was the reliance on student samples. Beyond general concerns about the representativeness of students, it is also likely that students' schedules differ in systematic ways from the general population, and these differences in time use could affect our results. However, it is not clear whether the differences in time use would necessarily affect cross-method agreement. Indeed, an argument could be made that there may be *more* variance in activities among students than among the general population because their time may be less stringently scheduled. Of course, future replications with non-student samples would be needed to test this possibility.

A third limitation is that although the lack of correspondence between the ESM and DRM leads to questions about the validity of one or both methods, the design of our study does not allow us to draw strong conclusions about whether one of the two methods has greater validity than the other. In our studies, we were able to provide some initial comparisons between both the ESM and DRM and alternative methods including informant reports, however these comparisons were inconclusive in regard to relative validity. Thus, future research should incorporate features that might add to knowledge about this critical issue. For instance, studies that have additional non-self-report measures that are assessed on a moment-to-moment basis (e.g., Bylsma et al., 2011) or studies that use aggregated experiential measures to predict future outcomes (including health or behavioral outcomes) could help in this regard.

Finally, given the nature of the ESM and DRM procedures, it was necessary in both studies to first assess experiences with the ESM and then to have participants retrospectively report about these experiences using the DRM. In other words, the ESM always preceded the DRM. This design feature leaves open the possibility that taking part in an ESM study prompts participants to pay more attention to their momentary experiences than they normally would, a feature that may make the DRM more accurate than it would be had it not followed participation in ESM. This could lead to an exaggeration of the extent to which the two methods agree. This concern is difficult to address with simple modifications of the design used in our studies. For sample-level statistics, it would be reasonable to assess the

DRM on one day and compare it to results from an ESM assessment on a separate, but similar day. However, even for person-level statistics—but especially for within-person estimates—assessing different days makes cross-method comparisons difficult. Arguably, however, the type of accuracy that might be most affected by this type of increased attention to moment-to-moment experience would be within-person accuracy, which in our study was quite low. Thus, if these accuracy-enhancing effects are at work, they do not seem to be resulting in high levels of accuracy at the within-person level. Regardless, we encourage researchers to examine these questions with more sophisticated designs in future work.

### **Theoretical Implications: Tentative Findings Regarding the Differences Between Methods**

Although an important goal for these studies was to assess the simple agreement between the two methods, we also sought to understand how results of more substantive questions may differ depending on the method that was used. Responses to ESM-based surveys likely reflect experiential knowledge, whereas participants may rely on episodic memory and situation-specific beliefs that lead to deviations from actual experience when completing the DRM (Robinson & Clore, 2002). If so, the associations between affect and various other predictors may differ depending on the method that is used.

In Study 1, we tested this possibility by examining the interaction between the method of assessment (ESM or DRM) and each situation variable, when predicting affective outcomes. The general pattern of results that emerged was that participants tended to report more extreme scores for positive affect when using the DRM as compared to the ESM. In other words, situations that were typically experienced as positive by participants were reported as being more strongly positive when using the DRM as compared to the ESM; and situations that were typically experienced as negative were reported as being less positive when using the DRM as compared to the ESM. This means that substantive conclusions about the role of situations could differ depending on the method that was used. The specific finding is consistent with the idea that situational factors play a bigger role in the DRM than ESM. In Study 2, this pattern again emerged, though the effects were significant for fewer individual situation variables than in Study 1.

It is important to note that this interaction between method and situation when predicting affect was much more apparent for positive than negative affect; a finding that replicated across both studies. Although it is not clear why this difference emerged, one possibility is that this discrepancy results from the simple fact that people typically report lower levels of negative affect, and hence, have lower levels of variability. With reduced variability in negative affect, method-specific effects may be subtler and more difficult to detect. Future research (perhaps with a specific focus on sampling more intensely negative situational factors) will be needed to address this possibility.

Studies 1 and 2 also included additional criterion variables that could be used to assess the substantive differences between responses to ESM and DRM; and again, these results provided some tentative evidence that DRM-based measures may be more strongly influenced by situations, or even situation expectations than ESM. For instance, in both studies, DRM measures were less strongly related to global measures of well-being than were ESM measures, including both informant and self-rated well-being. In contrast,

expectations regarding the typical affective experiences that one has in specific situations were somewhat more strongly correlated with DRM-based affect ratings than with ESM. Again, it is important to emphasize that these underpowered analyses were not significant, and thus, any conclusions about these differences must be extremely tentative. However, the pattern, when considered in light of the method-by-situation interactions and the differences in associations with global measures of well-being provides suggestive evidence that participants rely more on situational factors when providing DRM reports than when responding to ESM.

This is an important finding because it leads to questions about how people are responding to DRM-based measures, and perhaps experiential measures in general. One frequent finding from well-being research that uses the DRM is that stable, cross-situationally consistent factors correlate less strongly with experiential measures than with global measures, whereas the reverse is true for situational factors that occur at the time of the affective experience. Some researchers conclude, based on this evidence, that the associations with global ratings are based on people's *beliefs* about the factors that affect well-being, whereas the associations with more experiential measures reflect a more accurate picture of the impact of moment-to-moment situational factors on actual affective experience (Kahneman, 1999; Kahneman & Riis, 2005). The results presented here suggest a possible alternative interpretation of these results.

Some have argued that experiential measures may be *easier* for participants to answer, precisely because these measures do not require respondents to rely on memory for events or to aggregate across multiple experiences (Kahneman, 1999). However, experiential measures like the DRM and ESM are similar to global measures in that respondents must interpret the question to answer it. In some cases, respondents can infer more than the researchers intended, which can affect the answers that they provide (Schwarz, 1999). Most relevant to the issues in this paper, Baird and Lucas (2011) examined how ratings of role-specific personality varied depending on whether the fact that cross-role variance was a focus or not. Specifically, they asked participants to report about their personality in specific roles, experimentally manipulating whether participants were asked about their personality in a single role or across multiple roles. Baird and Lucas (2011) hypothesized that simply asking about multiple roles makes salient that the researchers are interested in the *differences* across roles, which in turn, prompts respondents to highlight these differences. In accordance with these predictions, cross-role variability was higher in the multiple role condition and correlations with global ratings of personality were weaker in this condition as compared to a condition where respondents only reported on a single role. In other words, when asked about a single role, respondents provided ratings that were quite consistent with their responses to a question about how their personality was in general; when asked about multiple roles, their general personality played less of a role and the impact of roles was emphasized.

It is possible that the DRM—or even experiential measures more generally—have the same effect as the multiple-role manipulation in Baird and Lucas (2011). By repeatedly asking about affect in different situations, researchers may implicitly convey to respondents that they are especially interested in *change* rather than stability. This effect may be especially

impactful in the DRM as compared to the ESM because the various “episodes” that participants must describe are presented one after another in the same setting (just as Baird and Lucas (2011) presented their role-specific questions one after another). This explanation would be consistent with all three effects reported above: That situational factors are more strongly associated with the DRM than the ESM, that DRM reports are less strongly associated with trait measures than are ESM reports, and the tentative finding that DRM reports are more closely linked with expectations for situations.

If this interpretation is correct, then this would have implications for our understanding of the relative importance of situational factors when predicting global versus experiential affect. As noted above, when experiential measures of well-being are assessed, the activities in which people participated seem to have the biggest effect on the well-being that they report, whereas broader, person-level factors, such as age, income, and health are only weakly related to these experiential measures (Kahneman et al., 2004). In contrast, when global well-being measures are examined, these same person-level factors seem to play a larger role. Because experiential measures are thought by some to “solve” the memory and aggregation problems inherent in global reports—features that makes these experiential measures more “objective” (Kahneman, 1999)—this discrepancy is often interpreted to mean that experiential measures like the DRM are more valid than global reports. The tentative conclusions that we draw from our study suggest that the DRM and perhaps other experiential measures may have their own unique problems that result from drawing attention to the situational factors in which one is engaged at the time the affective reports are provided.

Importantly, these types of effects have implications beyond the search for predictors of subjective well-being. For instance, Srivastava et al. (2008) and Oerlemans and Bakker (2014b) both used the DRM to assess extraverts’ and introverts experiences over the course of the day. One question they addressed was whether extraverts get a bigger “boost” in positive affect from social situations as compared to introverts (see Lucas, Le, & Dyrenforth, 2008, for a similar investigation using ESM), a question that is important for understanding the nature of extraversion. If experiential measures like the ESM and DRM enhance the salience of the situational characteristics that respondents experience, then this may also prompt these respondents to inflate the impact of these situational features in ways that are consistent with their own self views, including their self-perceived levels of traits like extraversion. Again, it must be emphasized that although this is an important potential implication of these findings, future research can clarify these effects by borrowing the methodology of Baird and Lucas (2011) or through further investigations of the role of expectations in DRM-based judgments.

## Conclusion

A substantial body of research shows that different methods for assessing subjective well-being do not perfectly cohere. One major focus of research and theory has been on the discrepancies between global and experiential measures (Kahneman et al., 2004; Kahneman & Riis, 2005). Because global measures require respondents to remember and aggregate across multiple experiences, some have suggested that they are prone to biases that affect



their validity (Kahneman, 1999; Schwarz & Strack, 1999). Experiential measures have been proposed as a useful, and perhaps even more valid, alternative because they do not rely on memory and aggregation. Yet at the same time, experiential measures are more time intensive than global measures, which has led to a search for less burdensome alternatives such as the DRM. The goal of our studies were first to investigate agreement between the “gold-standard” experiential measure, ESM, and a shorter, less intensive alternative, the DRM. We then hoped to use that information to develop a better understanding of the strengths and weaknesses of these measures, along with the processes that underlie them.

The clearest conclusion from our studies was that the correspondence between ESM-based and DRM-based measures of affective experience varies dramatically, depending on one’s goals. Aggregate, cross-person ratings of situational characteristics—the amount of time people spend in specific situations and the affect that they typically experience in them—cohere quite well across the two methods. However, as one moves to less and less aggregated levels, cross-method agreement declines. Although agreement about between-person differences in aggregated affect is quite strong, agreement about between-person differences in time spent in situations is only moderate and it varies across situations. Within-person agreement is quite low, both for situations and affect. This suggests that for many common purposes, it is not safe to assume that the DRM provides the same information as ESM.

Our studies also looked more closely at the nature of the differences across the two methods, showing that DRM-based measures are less strongly associated with global ratings of affect (both self- and informant-reports), more closely linked with variability in situations, and perhaps more closely associated with expectations. These findings lead to potential concerns that the relatively strong link between situational factors and DRM-rated affect may result from features of the DRM itself. Thus, we caution researchers to interpret these differences in associations carefully, because they could result from specific methodological characteristics of the DRM.

Although these results raise questions about the validity and utility of the DRM, they also provide clear directions for future research. Our preliminary findings strongly suggest that more work on how expectations regarding the affective impact of specific situations actually do impact ratings of these situations in the context of the DRM is needed. In addition, more work that specifically focuses on how the context of the DRM shapes the answers that people provide can help clarify the processes that underlie both experiential and global measures of well-being. The ultimate benefit of this work will not only be methodological, but also theoretical, as it will be necessary to consider any theoretical work on the predictors of well-being in the context of the specific measures used to assess this critical outcome.

## Acknowledgments

All materials and code can be downloaded at: <https://osf.io/dg63b/>. Because of privacy concerns regarding the identifiability of data due to low-frequency events in the experience sampling and day reconstruction components, data are not posted publicly. However, Richard Lucas and Brent Donnellan have archived copies of the data that can be shared with researchers who sign an agreement not to share the data further. This research supported by National Institute on Aging Grant AG040715 awarded to Richard E. Lucas and Brent Donnellan. Carol Wallsworth is now at Amazon, Seattle, Washington. This work was completed prior to joining Amazon.

## References

- Aust F, & Barth M (2018). papaja: Create APA manuscripts with R Markdown. Retrieved from <https://github.com/crsh/papaja>
- Baird BM, & Lucas RE (2011). “ ... And How About Now? ”: Effects of Item Redundancy on Contextualized Self-Reports of Personality. *Journal of Personality*, 79(5), 1081–1112. [PubMed: 21241305]
- Bates D, Mächler M, Bolker B, & Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. 10.18637/jss.v067.i01
- Bolger N, Davis A, & Rafaeli E (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616. 10.1146/annurev.psych.54.101601.145030
- Brown C (2012). Dummies: Create dummy indicator variables flexibly and efficiently. Retrieved from <https://CRAN.R-project.org/package=dummies>
- Bylsma LM, Taylor-Clift A, & Rottenberg J (2011). Emotional reactivity to daily events in major and minor depression. *Journal of Abnormal Psychology*, 120(1), 155–167. 10.1037/a0021662 [PubMed: 21319928]
- Daly M, Baumeister RF, Delaney L, & MacLachlan M (2014). Self-control and its relation to emotions and psychobiology: Evidence from a Day Reconstruction Method study. *Journal of Behavioral Medicine*, 27(1), 81–93. 10.1007/s10865-012-9470-9
- Daly M, Delaney L, Doran PP, Harmon C, & MacLachlan M (2010). Naturalistic monitoring of the affect-heart rate relationship: A day reconstruction study. *Health Psychology*, 29(2), 186–195. 10.1037/a0017626 [PubMed: 20230092]
- Diener E, Emmons RA, Larsen RJ, & Griffin S (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49, 71–75. [PubMed: 16367493]
- Diener E, & Tay L (2013). Review of the Day Reconstruction Method (DRM). *Social Indicators Research*, 116(1), 255–267. 10.1007/s11205-013-0279-x
- Dockray S, Grant N, Stone AA, Kahneman D, Wardle J, & Steptoe A (2010). A Comparison of Affect Ratings Obtained with Ecological Momentary Assessment and the Day Reconstruction Method. *Social Indicators Research*, 99(2), 269–283. 10.1007/s11205-010-9578-7 [PubMed: 21113328]
- Epstein S (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 27(1), 1097.
- Fox J, & Weisberg S (2019). *An R companion to applied regression (Third)*. Thousand Oaks CA: Sage. Retrieved from <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Gamer M, Lemon J, & <puspendra.pusp22@gmail.com>, I. F. P. S. (2019). Irr: Various coefficients of interrater reliability and agreement. Retrieved from <https://CRAN.R-project.org/package=irr>
- Harrell FE Jr, Charles Dupont, & others. (2019). Hmisc: Harrell miscellaneous. Retrieved from <https://CRAN.R-project.org/package=Hmisc>
- Hudson NW, Anusic I, Lucas RE, & Donnellan MB (2017). Comparing the Reliability and Validity of Global Self-Report Measures of Subjective Well-Being With Experiential Day Reconstruction Measures. *Assessment*, 1073191117744660. 10.1177/1073191117744660
- Hudson NW, Lucas RE, & Donnellan MB (2019). A direct comparison of the temporal stability and criterion validities of experiential and retrospective global measures of subjective well-being. Manuscript in Preparation, Dallas, TX: Southern Methodist University.
- Kahneman D (1999). Objective happiness. In Kahneman D, Diener E, & Schwarz N (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 3–25). New York, NY: Russell Sage Foundation.
- Kahneman D, Krueger AB, Schkade DA, Schwarz N, & Stone AA (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science (New York, N.Y.)*, 506(5702), 1776–1780. 10.1126/science.1103572
- Kahneman D, & Riis J (2005). Living, and thinking about it: Two perspectives on life. In Huppert FA, Baylis N, & Keveme B (Eds.), *The science of well-being* (pp. 285–304). New York: Oxford University Press, USA.

- Kim J, Kikuchi H, & Yamamoto Y (2013). Systematic comparison between ecological momentary assessment and day reconstruction method for fatigue and mood states in healthy adults. *British Journal of Health Psychology*, 75(1), 155–167. 10.1111/bjhp.12000
- Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). ImerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. 10.18637/jss.v082.i13
- Lee PH, Tse ACY, & Lee KY (2017). A new statistical model for the Day Reconstruction Method: Day Reconstruction Method. *International Journal of Methods Psychiatric Research*, 26(4), e1547. 10.1002/mpr.1547
- Lucas RE, Freedman VA, & Carr D (2019). Measuring experiential well-being among older adults. *The Journal of Positive Psychology*, 0(0), 1–10. 10.1080/17439760.2018.1497686
- Lucas RE, Le K, & Dyrenforth PS (2008). Explaining the extraversion/positive affect relation: Sociability cannot account for extraverts' greater happiness. *Journal of Personality*, 76(3), 385–414. [PubMed: 18399958]
- Oerlemans WGM, & Bakker AB (2014a). Burnout and daily recovery: A day reconstruction study. *Journal of Occupational Health Psychology*, 19(3), 303–314. 10.1037/a0036904 [PubMed: 24885683]
- Oerlemans WGM, & Bakker AB (2014b). Why extraverts are happier: A day reconstruction study. *Journal of Research in Personality*, 50, 11–22. 10.1016/j.jrp.2014.02.001
- Pinheiro J, Bates D, DebRoy S, Sarkar D, & R Core Team. (2019). nlme: Linear and nonlinear mixed effects models. Retrieved from <https://CRAN.R-project.org/package=nlme>
- R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Revelle W (2018). Psych: Procedures for psychological, psychometric, and personality research. Evanston, Illinois: Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- Rinker TW (2019). qdap: Quantitative discourse analysis package. Buffalo, New York. Retrieved from <http://github.com/trinker/qdap>
- Robinson MD, & Clore GL (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128(6), 934–960. 10.1037/0033-2909.128.6.934 [PubMed: 12405138]
- Schwarz N (1999). Self-Reports: How the Questions Shape the Answers. *American Psychologist*, 54(2), 93–105.
- Schwarz N, & Strack F (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In Kahneman D, Diener E, & Schwarz N (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York, NY: Russell Sage Foundation.
- Scollon CN, Howard AH, Caldwell AE, & Ito S (2009). The Role of Ideal Affect in the Experience and Memory of Emotions. *Journal of Happiness Studies*, 10(3), 257–269. 10.1007/s10902-007-9079-9
- Srivastava S, Angelo KM, & Vallereux SR (2008). Extraversion and positive affect: A day reconstruction study of person-environment transactions. *Journal of Research in Personality*, 42(6), 1613–1618. <https://doi.org/10/fw3s3c>
- Stone AA, Schwartz JE, Schkade D, Schwarz N, Krueger A, & Kahneman D (2006). A population approach to the study of emotion: Diurnal rhythms of a working day examined with the day reconstruction method. *Emotion*, 6(1), 139–149. 10.1037/1528-3542.6.1.139 [PubMed: 16637757]
- Tay L, Chan D, & Diener E (2013). The Metrics of Societal Happiness. *Social Indicators Research*, 117(5 2013), 577–600. 10.1007/s11205-013-0356-1
- Wickham H (2016). Ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham H, François R, Henry L, & Müller K (2019). Dplyr: A grammar of data manipulation. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham H, & Henry L (2019). Tidy: Tidy messy data. Retrieved from <https://CRAN.R-project.org/package=tidy>
- Wickham H, Hester J, & François R (2018). Readr: Read rectangular text data. Retrieved from <https://CRAN.R-project.org/package=readr>

Zeileis A, & Croissant Y (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, 24(1), 1–13. 10.18637/jss.v034.i01

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1:**

Sample-level descriptive statistics and paired t-tests for each emotion and each method in Study 1

Emotion	ESM	DRM	d	t	df	P
Happy	5.17(0.98)	4.94(1.13)	0.22	7.16	408	< 0.001
Frustrated	2.23(1.02)	2.10(1.03)	0.12	3.40	407	< 0.001
Sad	1.85(1.01)	1.71(1.01)	0.14	4.55	407	< 0.001
Satisfied	4.52(1.22)	4.27(1.38)	0.19	6.51	407	< 0.001
Angry	1.62(0.83)	1.55(0.87)	0.08	2.07	407	0.039
Worried	2.66(1.43)	2.41(1.39)	0.18	6.53	407	< 0.001
Tired	3.81(1.30)	3.72(1.40)	0.06	1.79	407	0.074
Pain	1.61(0.97)	1.57(0.95)	0.05	1.07	407	0.286
Meaning	3.83(1.61)	3.69(1.70)	0.08	3.77	407	< 0.001
Positive Affect	4.85(1.02)	4.61(1.16)	0.22	7.72	408	< 0.001
Negative Affect	2.09(0.93)	1.95(0.94)	0.16	5.40	407	< 0.001

*Note.* Total N = 425 for ESM data and 410 for DRM data; sample sizes are occasionally reduced due to nonresponse on specific items. Paired t-tests are based only on participants who have valid data from both methods of assessment and thus do not correspond exactly to the mean differences reported in Columns 1 through 3. Items were rated on a scale from 1 to 7. ESM = Experience Sampling Method; DRM = Day Reconstruction Method.

**Table 2:**

Sample-Level Descriptive Statistics for Situation Ratings from DRM and ESM in Study 1

Situation	DRM Time	DRM Episodes	ESM Episodes
What: Commuting	6.41 (9.89)	14.56 (15.86)	7.20 (10.37)
What: Errands	2.76 (8.01)	3.31 (8.75)	2.15 (5.67)
What: Housework	2.81 (7.22)	2.77 (6.55)	2.92 (6.93)
What: Appointment	1.09 (3.65)	1.47 (4.67)	0.88 (3.65)
What: Work	5.27 (11.97)	3.49 (7.79)	5.68 (13.20)
What: Class	13.46 (15.66)	13.42 (14.43)	13.42 (13.58)
What: Homework	17.86 (20.61)	13.01 (14.74)	18.39 (18.87)
What: FoodPrep	2.42 (6.49)	4.04 (7.99)	2.99 (6.89)
What: Eating	13.66 (16.04)	19.01 (14.28)	14.56 (13.16)
What: Socializing	22.09 (24.27)	20.01 (21.52)	14.93 (18.15)
What: Sex	1.17 (6.82)	1.02 (5.07)	1.00 (4.33)
What: Grooming	3.28 (8.04)	5.66 (8.06)	4.78 (7.89)
What: Sleeping	15.13 (19.67)	7.52 (8.52)	7.50 (11.12)
What: Relaxing	18.03 (21.72)	14.68 (16.77)	19.08 (18.36)
What: Reading	1.46 (5.41)	1.22 (4.36)	1.13 (4.84)
What: Entertaining	11.94 (17.74)	8.89 (13.65)	8.29 (14.13)
What: Exercise	2.20 (5.76)	2.84 (6.62)	2.70 (7.77)
What: TV	12.23 (18.28)	9.28 (13.50)	10.18 (14.37)
What: Internet	9.12 (15.48)	6.54 (11.11)	9.90 (14.92)
What: Phone	5.63 (14.75)	5.93 (13.95)	8.30 (15.99)
What: Worship	0.65 (4.85)	0.50 (2.75)	0.50 (3.26)
Who: Nobody	38.87 (28.15)	41.35 (24.67)	40.95 (23.92)
Who: Partner	8.37 (21.23)	7.33 (18.86)	7.44 (16.68)
Who: Friends	32.49 (28.29)	30.63 (25.00)	30.87 (24.50)
Who: Roommate	19.06 (25.77)	17.89 (22.23)	17.66 (22.22)
Who: Patrons	2.44 (8.70)	1.70 (6.03)	2.63 (8.50)
Who: Colleagues	14.76 (18.73)	12.68 (15.80)	11.74 (15.95)
Who: Boss	1.84 (7.16)	1.08 (4.28)	1.63 (6.63)
Who: Family	5.63 (15.55)	5.44 (14.68)	4.67 (12.46)
Who: Other	3.27 (8.96)	3.96 (8.90)	2.40 (6.19)
Where: Home	46.16 (30.44)	40.11 (26.70)	41.39 (27.54)
Where: School	28.84 (29.90)	32.71 (29.83)	35.30 (29.69)
Where: Work	4.34 (11.40)	3.02 (8.42)	5.46 (13.14)
Where: Other	19.84 (24.40)	23.06 (23.04)	17.65 (20.70)

*Note.* N for DRM variables = 410, N for ESM variables = 425, DRM = Day Reconstruction Method; ESM = Experience Sampling Method. Data from all moments were included, regardless of whether they came from moments for which both methods of assessment were available.



**Table 3:**

Sample-Level Situation-Specific Affect Ratings Based on DRM and ESM Ratings in Study 1

Situation	Positive Affect		Negative Affect	
	DRM	ESM	DRM	ESM
What: Commuting	4.24 (1.48)	4.76 (1.42)	1.98 (1.04)	2.23 (1.21)
What: Errands	5.06 (1.30)	5.40 (1.30)	1.92 (1.03)	2.06 (1.41)
What: Housework	4.62 (1.41)	4.97 (1.25)	2.06 (1.19)	2.23 (1.19)
What: Appointment	4.54 (1.39)	4.57 (1.38)	2.13 (1.20)	2.08 (1.48)
What: Work	4.21 (1.53)	4.68 (1.33)	1.95 (1.04)	1.98 (0.97)
What: Class	3.98 (1.30)	4.50(1.14)	2.15 (1.06)	2.29 (1.11)
What: Homework	4.04(1.41)	4.51 (1.20)	2.31 (1.14)	2.38 (1.09)
What: FoodPrep	4.88 (1.35)	4.93 (1.34)	1.73 (0.89)	1.92 (1.03)
What: Eating	5.18 (1.18)	5.23 (1.19)	1.75 (0.95)	1.98 (1.07)
What: Socializing	5.28 (1.26)	5.46 (1.16)	1.79 (0.94)	1.82 (1.00)
What: Sex	6.03 (1.28)	6.19 (0.92)	1.55 (0.83)	1.73 (1.18)
What: Grooming	4.78 (1.41)	4.85 (1.30)	1.77 (0.99)	2.03 (1.03)
What: Sleeping	4.99 (1.66)	4.49 (1.29)	1.73 (1.15)	2.17 (1.22)
What: Relaxing	4.97 (1.45)	5.01 (1.26)	1.90 (1.12)	1.98 (1.08)
What: Reading	5.28 (1.41)	5.09 (1.36)	1.79 (0.93)	1.93 (0.95)
What: Entertaining	5.40 (1.36)	5.26 (1.21)	1.66 (0.99)	1.83 (0.93)
What: Exercise	5.21 (1.30)	5.27 (1.24)	1.62 (0.93)	1.67 (0.95)
What: TV	5.09 (1.40)	5.12 (1.24)	1.75 (1.07)	1.89 (1.00)
What: Internet	4.69 (1.43)	4.77 (1.30)	1.95 (1.13)	2.11 (1.11)
What: Phone	4.61 (1.51)	4.81 (1.17)	2.00 (1.15)	2.26 (1.23)
What: Worship	6.09 (1.02)	5.60 (1.11)	1.46 (0.49)	1.80 (0.90)
Who: Nobody	4.39 (1.33)	4.59 (1.12)	1.96 (1.03)	2.20 (1.06)
Who: Partner	4.99 (1.36)	5.43 (1.30)	1.85 (0.97)	1.90 (1.03)
Who: Friends	4.94 (1.24)	5.15 (1.10)	1.90 (1.00)	2.00 (0.99)
Who: Roommate	4.90 (1.31)	5.04 (1.20)	1.85 (1.02)	2.03 (1.14)
Who: Patrons	4.41 (1.32)	4.83 (1.20)	1.99 (0.99)	1.94 (0.95)
Who: Colleagues	4.17 (1.35)	4.71 (1.18)	2.14 (1.05)	2.16 (0.99)
Who: Boss	4.23 (1.44)	4.78 (1.37)	1.99 (0.89)	1.89 (0.78)
Who: Family	5.22 (1.45)	5.51 (1.21)	1.86 (1.06)	1.79 (0.91)
Who: Other	4.64 (1.54)	5.08 (1.53)	2.16 (1.26)	2.07 (1.15)
Where: Home	4.80 (1.25)	4.90 (1.12)	1.83 (0.95)	2.04 (1.04)
Where: School	4.17 (1.28)	4.65 (1.14)	2.09 (1.03)	2.22 (1.08)
Where: Work	4.13 (1.48)	4.74 (1.37)	1.94 (0.93)	1.96 (0.99)
Where: Other	4.87 (1.43)	5.17 (1.28)	1.85 (1.05)	1.95 (1.05)

*Note.* DRM = Day Reconstruction Method; ESM = Experience Sampling Method; Ns vary by situation variable, depending on how many people participated in that situation. Data from all moments were included, regardless of whether they came from moments for which both methods of assessment were available (i.e., “matched” moments).

**Table 4:**

Between- and within-person correlations between ESM and DRM-reported emotions from Study 1

<b>Emotion</b>	<b>Between r</b>	<b>Within r</b>
Happy	0.81	0.37 (0.50)
Frustrated	0.76	0.33 (0.55)
Sad	0.86	0.19 (0.56)
Satisfied	0.81	0.21 (0.52)
Angry	0.80	0.33 (0.54)
Worried	0.86	0.27 (0.53)
Tired	0.80	0.40 (0.53)
Pain	0.81	0.27 (0.55)
Meaning	0.90	0.20 (0.57)
Positive Affect	0.83	0.34 (0.54)
Negative Affect	0.88	0.36 (0.54)

*Note.* Overall N = 401, though sample sizes for within-person correlations vary because of missing data. Between-person standard deviations of within-person correlations are reported in parentheses in Column 2.

**Table 5:**

Cross-Method Agreement (Between- and Within-Person) for Situation Ratings in Study 1

Situation	Between-Person r		Within-Person kappa	No Variance
	DRM Time	DRM Episodes		
What: Commuting	0.25	0.35	0.27 (0.42)	56%
What: Errands	0.57	0.56	0.29 (0.46)	82%
What: Housework	0.28	0.34	0.18 (0.38)	79%
What: Appointment	0.44	0.45	0.22 (0.40)	90%
What: Work	0.81	0.70	0.50 (0.48)	76%
What: Class	0.68	0.67	0.72 (0.42)	42%
What: Homework	0.59	0.60	0.41 (0.44)	36%
What: FoodPrep	0.20	0.26	0.22 (0.43)	79%
What: Eating	0.21	0.32	0.31 (0.47)	29%
What: Socializing	0.49	0.43	0.27 (0.43)	36%
What: Sex	0.26	0.40	0.20 (0.41)	92%
What: Grooming	0.17	0.31	0.20 (0.39)	76%
What: Sleeping	0.41	0.38	0.26 (0.42)	74%
What: Relaxing	0.44	0.47	0.21 (0.41)	36%
What: Reading	0.38	0.36	0.20 (0.38)	91%
What: Entertaining	0.58	0.56	0.24 (0.41)	58%
What: Exercise	0.29	0.36	0.36 (0.48)	81%
What: TV	0.51	0.54	0.26 (0.43)	53%
What: Internet	0.52	0.51	0.17 (0.40)	57%
What: Phone	0.46	0.45	0.05 (0.27)	69%
What: Worship	0.28	0.45	0.36 (0.52)	94%
Who: Nobody	0.53	0.54	0.37 (0.44)	15%
Who: Partner	0.74	0.74	0.40 (0.46)	78%
Who: Friends	0.58	0.56	0.40 (0.44)	22%
Who: Roommate	0.68	0.65	0.34 (0.47)	47%
Who: Patrons	0.68	0.65	0.38 (0.47)	86%
Who: Colleagues	0.58	0.55	0.47 (0.46)	47%
Who: Boss	0.76	0.74	0.48 (0.50)	90%
Who: Family	0.71	0.60	0.29 (0.42)	79%
Who: Other	0.42	0.49	0.19 (0.40)	78%
Where: Home	0.77	0.68	0.61 (0.41)	26%
Where: School	0.79	0.70	0.64 (0.42)	34%
Where: Work	0.83	0.76	0.60 (0.46)	81%
Where: Other	0.74	0.67	0.45 (0.45)	40%

*Note.* Total N = 401, though sample sizes vary across items due to missing data or lack of variance. The column labelled “No Variance” reflects the percentage of respondents who had zero variance on one or more more methods for a specific situation, resulting in kappa not being computed for that person. Between-person standard deviations in within-person kappa are reported in parentheses in Column 3.

**Table 6:** Interactions Between Method and Situation Variables When Predicting Affect in Study 1

Situation	Positive Affect					Negative Affect				
	Estimate	SE	df	t	p	Estimate	SE	df	t	p
What: Commuting	-0.17	0.05	3613	-3.05	0.002	-0.01	0.04	3608	-0.20	0.845
What: Errands	-0.02	0.09	3613	-0.22	0.826	-0.01	0.06	3608	-0.16	0.873
What: Housework	-0.03	0.09	3613	-0.29	0.770	0.02	0.06	3608	0.33	0.744
What: Appointment	-0.19	0.15	3613	-1.30	0.192	0.20	0.10	3608	1.98	0.048
What: Work	-0.07	0.06	3613	-1.09	0.278	0.03	0.04	3608	0.74	0.459
What: Class	-0.19	0.04	3613	-4.78	<0.001	0.03	0.03	3608	1.26	0.207
What: Homework	-0.12	0.04	3613	-3.17	0.002	0.04	0.02	3608	1.77	0.077
What: FoodPrep	0.17	0.09	3613	1.94	0.053	-0.10	0.06	3608	-1.69	0.092
What: Eating	0.15	0.04	3613	3.82	<0.001	-0.05	0.03	3608	-1.71	0.088
What: Socializing	0.12	0.04	3613	3.00	0.003	0.03	0.03	3608	0.98	0.328
What: Sex	0.06	0.16	3613	0.35	0.728	0.04	0.11	3608	0.37	0.712
What: Grooming	0.02	0.09	3613	0.24	0.811	-0.08	0.06	3608	-1.34	0.179
What: Sleeping	0.39	0.08	3613	5.18	<0.001	-0.21	0.05	3608	-4.18	<0.001
What: Relaxing	0.16	0.04	3613	3.78	<0.001	-0.05	0.03	3608	-1.81	0.070
What: Reading	0.13	0.14	3613	0.93	0.351	0.03	0.10	3608	0.26	0.794
What: Entertaining	0.17	0.05	3613	3.39	<0.001	0.00	0.03	3608	0.02	0.987
What: Exercise	0.10	0.09	3613	1.09	0.276	0.04	0.06	3608	0.59	0.554
What: TV	0.17	0.05	3613	3.61	<0.001	-0.04	0.03	3608	-1.22	0.223
What: Internet	0.01	0.05	3613	0.27	0.790	0.00	0.03	3608	0.03	0.973
What: Phone	0.06	0.06	3613	0.93	0.351	-0.06	0.04	3608	-1.41	0.157
What: Worship	0.18	0.18	3613	1.03	0.303	-0.17	0.12	3608	-1.42	0.155
Who: Nobody	-0.05	0.03	3613	-1.44	0.150	-0.03	0.02	3608	-1.23	0.217
Who: Partner	0.04	0.06	3613	0.66	0.509	-0.06	0.04	3608	-1.48	0.139
Who: Friends	0.07	0.03	3613	2.33	0.020	0.02	0.02	3608	0.97	0.334
Who: Roommate	0.14	0.04	3613	3.56	<0.001	-0.01	0.03	3608	-0.38	0.706
Who: Patrons	-0.19	0.09	3613	-2.23	0.026	0.18	0.06	3608	3.16	0.002
Who: Colleagues	-0.12	0.04	3613	-2.88	0.004	0.05	0.03	3608	1.82	0.069

Situation	Positive Affect				Negative Affect				
	Estimate	SE	df	t	Estimate	SE	df	t	p
Who: Boss	-0.19	0.10	3613	-1.83	0.067	0.07	3608	1.31	0.189
Who: Family	0.10	0.07	3613	1.48	0.139	0.05	3608	-0.05	0.962
Who: Other	-0.15	0.09	3613	-1.68	0.093	0.06	3608	1.01	0.311
Where: Home	0.08	0.03	3613	2.70	0.007	0.02	3608	0.17	0.867
Where: School	-0.13	0.03	3613	-4.13	<0.001	0.01	3608	0.34	0.733
Where: Work	-0.08	0.07	3613	-1.24	0.214	0.03	3608	0.72	0.469
Where: Other	0.08	0.04	3613	2.08	0.038	-0.02	3608	-0.93	0.352

Note. N = 401.

**Table 7:**

Between-person correlations between global and informant measures of well-being and experiential measures in Study 1

	Self Report				Informant Report	
	Life Satisfaction	SWLS	Global PA	Global NA	Life Satisfaction	SWLS
ESM PA	0.49	0.47	0.53	-0.28	0.27	0.27
ESM NA	-0.34	-0.27	-0.34	0.41	-0.12	-0.12
DRM PA	0.40	0.38	0.42	-0.24	0.26	0.24
DRM NA	-0.33	-0.24	-0.33	0.39	-0.08	-0.08

*Note.* Ns range from 303 to 409. DRM = Day Reconstruction Method; ESM = Experience Sampling Method; PA = Positive Affect; NA = Negative affect; SWLS = Satisfaction With Life Scale.



**Table 8:**

Sample-level descriptive statistics and paired t-tests for each emotion and each method in Study 2

Emotion	ESM	DRM	d	t	df	p
Happy	5.06(0.90)	5.06(0.93)	0.00	0.08	418	0.938
Frustrated	2.37(1.10)	2.21(0.97)	0.15	5.04	418	< 0.001
Sad	1.94(0.94)	1.82(0.91)	0.13	5.01	418	< 0.001
Satisfied	4.63(1.01)	4.61(1.00)	0.02	-0.25	418	0.806
Angry	1.79(0.89)	1.75(0.86)	0.05	2.04	418	0.042
Worried	2.80(1.44)	2.52(1.29)	0.20	7.57	418	< 0.001
Tired	4.04(1.36)	3.95(1.28)	0.07	1.93	418	0.054
Pain	1.67(0.96)	1.64(0.92)	0.03	1.06	418	0.288
Meaning	4.36(1.27)	4.21(1.33)	0.11	3.61	418	< 0.001
Positive Affect	4.85(0.90)	4.84(0.90)	0.01	-0.10	418	0.920
Negative Affect	2.22(0.94)	2.08(0.90)	0.16	6.52	418	< 0.001

*Note.* N = 419 for ESM, N = 461 for DRM. Paired t-tests only include participants who provided both ESM and DRM data and thus do not correspond exactly to the mean differences reported in Columns 1 through 3. Items were rated on a scale from 1 to 7.

**Table 9:**

Sample-level descriptive statistics for situation ratings for DRM and ESM in Study 2

Situation	DRM Time	DRM Episodes	ESM Episodes
What: Commuting	5.92 (7.65)	14.94 (13.20)	6.89 (11.87)
What: Errands	0.90 (2.91)	1.45 (3.86)	2.00 (7.10)
What: Housework	2.47 (5.58)	2.96 (5.97)	4.28 (10.03)
What: Appointment	0.67 (2.30)	1.01 (3.05)	0.85 (3.75)
What: Work	3.26 (7.87)	2.64 (5.69)	4.84 (12.28)
What: Class	11.65 (8.84)	12.92 (9.28)	19.92 (17.90)
What: Homework	13.67 (11.80)	13.51 (10.96)	19.97 (20.45)
What: FoodPrep	2.47 (6.08)	3.64 (7.35)	1.90 (7.06)
What: Eating	10.33 (9.26)	16.19 (9.64)	11.73 (13.42)
What: Socializing	9.82 (12.38)	12.00 (14.64)	11.56 (17.07)
What: Sex	0.91 (4.36)	0.95 (3.37)	1.27 (5.90)
What: Grooming	4.43 (4.98)	9.09 (7.73)	4.91 (9.01)
What: Sleeping	33.70 (12.68)	17.50 (11.01)	4.85 (9.12)
What: Relaxing	12.18 (13.68)	14.30 (13.33)	18.36 (20.56)
What: Reading	0.56 (2.68)	0.81 (3.50)	0.49 (2.70)
What: Entertaining	6.50 (9.53)	7.09 (10.03)	7.63 (14.53)
What: Exercise	2.50 (5.06)	3.67 (6.48)	2.81 (7.41)
What: TV	6.85 (9.96)	7.72 (10.39)	11.10 (18.44)
What: Internet	5.94 (10.81)	6.10 (10.40)	9.19 (16.76)
What: Phone	4.79 (10.22)	5.90 (11.36)	9.20 (18.55)
What: Worship	0.66 (3.25)	0.84 (3.64)	0.64 (3.93)
Who: Nobody	48.78 (23.63)	47.52 (20.59)	40.86 (24.06)
Who: Partner	7.08 (17.94)	7.68 (17.09)	7.68 (18.59)
Who: Friends	19.58 (16.95)	25.40 (20.20)	27.88 (25.23)
Who: Roommate	11.75 (15.72)	14.22 (17.17)	13.80 (20.23)
Who: Patrons	1.63 (5.51)	1.19 (3.88)	2.27 (8.95)
Who: Colleagues	11.80 (11.47)	12.54 (11.73)	13.89 (17.60)
Who: Boss	1.17 (5.14)	0.92 (3.62)	1.30 (7.37)
Who: Family	3.38 (10.10)	4.53 (12.11)	4.13 (12.45)
Who: Other	1.92 (5.07)	2.90 (6.83)	3.31 (9.78)
Where: Home	51.27 (27.33)	46.75 (25.50)	42.09 (28.50)
Where: School	28.17 (28.21)	35.40 (29.73)	39.82 (30.26)
Where: Work	2.56 (7.34)	1.95 (5.09)	4.28 (12.37)
Where: Other	9.74 (13.51)	15.85 (17.99)	13.80 (17.99)

Note. DRM = Day Reconstruction Method; ESM = Experience Sampling Method. N for DRM variables = 461; N for ESM variables = 419.

**Table 10:**

Sample-Level Situation-Specific Affect Ratings Based on DRM and ESM Ratings in Study 2

Situation	Positive Affect		Negative Affect	
	DRM	ESM	DRM	ESM
What: Commuting	4.42 (1.14)	4.63 (1.23)	2.18 (1.08)	2.34 (1.18)
What: Errands	4.98 (1.34)	5.00 (1.28)	2.20 (1.40)	2.03 (1.04)
What: Housework	4.63 (1.18)	4.73 (1.09)	2.26 (1.31)	2.53 (1.16)
What: Appointment	4.41 (1.64)	4.55 (1.23)	2.27 (1.15)	2.59 (1.15)
What: Work	4.49 (1.30)	4.65 (1.16)	2.33 (1.13)	2.17 (0.95)
What: Class	4.05 (1.21)	4.45 (1.09)	2.43 (1.26)	2.41 (1.18)
What: Homework	4.17 (1.24)	4.51 (1.11)	2.62 (1.26)	2.48 (1.13)
What: FoodPrep	5.00 (1.29)	4.98 (1.25)	1.96 (1.01)	2.22 (1.08)
What: Eating	5.31 (1.09)	5.33 (1.10)	1.88 (1.01)	2.06 (1.10)
What: Socializing	5.45 (1.12)	5.26 (1.08)	1.87 (0.99)	1.96 (0.96)
What: Sex	6.18 (1.00)	5.93 (1.13)	1.47 (0.70)	1.92 (1.09)
What: Grooming	4.57 (1.20)	4.95 (1.25)	2.05 (1.08)	2.10 (1.14)
What: Sleeping	5.41 (1.44)	4.92 (1.12)	1.72 (1.02)	2.16 (1.12)
What: Relaxing	5.26 (1.17)	5.20 (1.07)	1.94 (0.98)	2.03 (1.02)
What: Reading	5.34 (1.15)	5.61 (0.94)	2.13 (1.14)	1.55 (0.79)
What: Entertaining	5.46 (1.07)	5.25 (1.05)	1.87 (0.98)	1.94 (0.99)
What: Exercise	5.54 (1.20)	5.50 (1.04)	1.92 (0.98)	2.13 (0.98)
What: TV	5.32 (1.13)	5.19 (1.11)	1.87 (1.03)	1.97 (1.07)
What: Internet	4.73 (1.34)	4.63 (1.08)	2.16 (1.19)	2.20 (1.07)
What: Phone	4.76 (1.26)	4.89 (1.14)	2.25 (1.25)	2.23 (1.17)
What: Worship	5.27 (1.61)	5.31 (1.13)	1.97 (1.33)	2.23 (1.29)
Who: Nobody	4.78 (1.07)	4.62 (1.01)	2.07 (1.01)	2.30 (1.06)
Who: Partner	5.38 (1.25)	5.43 (1.17)	1.98 (1.02)	2.06 (1.05)
Who: Friends	5.06 (1.12)	5.16 (1.02)	2.03 (1.01)	2.11 (1.00)
Who: Roommate	4.93 (1.14)	4.95 (1.07)	2.11 (1.10)	2.18 (1.10)
Who: Patrons	4.73 (1.41)	4.75 (1.35)	2.30 (1.12)	2.32 (1.31)
Who: Colleagues	4.22 (1.28)	4.52 (1.15)	2.36 (1.25)	2.27 (1.10)
Who: Boss	4.81 (1.35)	4.97 (1.46)	2.08 (0.96)	2.11 (0.85)
Who: Family	5.22 (1.28)	5.06 (1.19)	2.02 (1.16)	2.23 (1.09)
Who: Other	4.71 (1.35)	4.69 (1.24)	2.21 (1.29)	2.46 (1.40)
Where: Home	5.01 (1.04)	4.97 (1.02)	2.00 (0.92)	2.17 (1.06)
Where: School	4.40 (1.12)	4.54 (1.04)	2.28 (1.10)	2.37 (1.11)
Where: Work	4.62 (1.29)	4.77 (1.13)	2.30 (1.09)	2.10 (1.01)
Where: Other	5.04 (1.10)	5.11 (1.21)	1.99 (0.97)	2.18 (1.12)

*Note.* DRM = Day Reconstruction Method; ESM = Experience Sampling Method. Ns vary across situations depending on how many participants reported participating in each activity.

**Table 11:**

Between- and within-person correlations between ESM and DRM-reported emotions from Study 2

<b>Emotion</b>	<b>Between r</b>	<b>Within r</b>
Happy	0.73	0.31 (0.48)
Frustrated	0.77	0.28 (0.50)
Sad	0.75	0.23 (0.47)
Satisfied	0.74	0.22 (0.48)
Angry	0.76	0.26 (0.51)
Worried	0.84	0.24 (0.47)
Tired	0.75	0.32 (0.47)
Pain	0.80	0.17 (0.51)
Meaning	0.88	0.19 (0.46)
Positive Affect	0.78	0.31 (0.48)
Negative Affect	0.84	0.33 (0.50)

*Note.* Overall N = 419, though sample sizes for within-person correlations vary because of missing data. Between-person standard deviations of within-person correlations are reported in parentheses in Column 2.

**Table 12:**

Cross-Method Agreement (Between- and Within-Person) for Situation Ratings in Study 2

Situation	Between-Person r		Within-Person kappa	No Variance
	DRM Time	DRM Episodes		
What: Commuting	0.17	0.23	0.17 (0.36)	51%
What: Errands	0.26	0.35	0.17 (0.36)	85%
What: Housework	0.33	0.33	0.11 (0.33)	72%
What: Appointment	0.30	0.42	0.24 (0.42)	90%
What: Work	0.83	0.84	0.53 (0.45)	79%
What: Class	0.65	0.60	0.64 (0.40)	24%
What: Homework	0.53	0.51	0.31 (0.41)	21%
What: FoodPrep	0.50	0.35	0.09 (0.30)	81%
What: Eating	0.28	0.31	0.24 (0.42)	23%
What: Socializing	0.55	0.56	0.21 (0.38)	43%
What: Sex	0.25	0.28	0.27 (0.44)	91%
What: Grooming	0.28	0.31	0.25 (0.43)	57%
What: Sleeping	0.29	0.29	0.26 (0.43)	58%
What: Relaxing	0.45	0.41	0.21 (0.38)	28%
What: Reading	0.37	0.40	0.11 (0.34)	94%
What: Entertaining	0.43	0.37	0.16 (0.34)	52%
What: Exercise	0.49	0.52	0.30 (0.45)	77%
What: TV	0.61	0.56	0.29 (0.39)	53%
What: Internet	0.56	0.48	0.12 (0.32)	59%
What: Phone	0.55	0.44	0.05 (0.30)	61%
What: Worship	0.12	0.10	0.04 (0.20)	93%
Who: Nobody	0.57	0.60	0.31 (0.40)	8%
Who: Partner	0.75	0.77	0.33 (0.44)	76%
Who: Friends	0.63	0.61	0.35 (0.42)	20%
Who: Roommate	0.64	0.62	0.24 (0.41)	44%
Who: Patrons	0.77	0.69	0.39 (0.45)	89%
Who: Colleagues	0.60	0.63	0.42 (0.43)	34%
Who: Boss	0.75	0.64	0.38 (0.41)	93%
Who: Family	0.80	0.76	0.30 (0.40)	80%
Who: Other	0.15	0.18	0.12 (0.32)	76%
Where: Home	0.72	0.68	0.56 (0.39)	19%
Where: School	0.77	0.75	0.58 (0.42)	21%
Where: Work	0.86	0.83	0.59 (0.46)	84%
Where: Other	0.70	0.75	0.37 (0.42)	42%

Note. N = 419, though sample sizes vary across items due to missing data or lack of variance.

The column labelled "No Variance" reflects the percentage of respondents who had zero variance in one more more methods for a specific situation, resulting in kappa not being computed for that person. Between-person standard deviations in within-person kappa are reported in parentheses in Column 3.

**Table 13:** Study 2 Interactions Between Method and Situation Variables When Predicting Affect

Situation	Positive Affect					Negative Affect				
	Estimate	SE	df	t	p	Estimate	SE	df	t	p
What: Commuting	-0.09	0.05	5013	-1.69	0.091	-0.04	0.04	5013	-1.00	0.319
What: Errands	0.19	0.10	5013	1.80	0.071	0.06	0.08	5013	0.75	0.454
What: Housework	-0.09	0.07	5013	-1.30	0.193	0.03	0.06	5013	0.56	0.575
What: Appointment	0.01	0.13	5013	0.11	0.915	-0.07	0.11	5013	-0.69	0.490
What: Work	-0.06	0.06	5013	-0.99	0.324	0.05	0.05	5013	1.12	0.264
What: Class	-0.21	0.03	5013	-6.54	<0.001	0.02	0.02	5013	0.65	0.514
What: Homework	-0.15	0.03	5013	-4.52	<0.001	0.12	0.03	5013	4.66	<0.001
What: FoodPrep	0.07	0.09	5013	0.78	0.435	-0.01	0.07	5013	-0.12	0.901
What: Eating	0.07	0.04	5013	1.82	0.068	-0.04	0.03	5013	-1.45	0.147
What: Socializing	0.15	0.04	5013	3.76	<0.001	-0.04	0.03	5013	-1.31	0.189
What: Sex	0.18	0.13	5013	1.39	0.164	-0.07	0.10	5013	-0.66	0.510
What: Grooming	-0.05	0.06	5013	-0.87	0.382	-0.07	0.05	5013	-1.60	0.109
What: Sleeping	0.38	0.06	5013	6.59	<0.001	-0.23	0.05	5013	-5.12	<0.001
What: Relaxing	0.15	0.04	5013	4.15	<0.001	-0.04	0.03	5013	-1.41	0.158
What: Reading	-0.16	0.17	5013	-0.95	0.344	0.20	0.13	5013	1.48	0.138
What: Entertaining	0.14	0.05	5013	2.91	0.004	0.00	0.04	5013	0.09	0.929
What: Exercise	0.01	0.08	5013	0.18	0.858	-0.04	0.06	5013	-0.63	0.530
What: TV	0.14	0.04	5013	3.32	<0.001	0.01	0.03	5013	0.17	0.863
What: Internet	0.04	0.05	5013	0.75	0.452	0.14	0.04	5013	3.51	<0.001
What: Phone	0.01	0.05	5013	0.16	0.870	0.11	0.04	5013	2.58	0.010
What: Worship	-0.15	0.16	5013	-0.92	0.356	0.11	0.13	5013	0.86	0.390
Who: Nobody	0.04	0.03	5013	1.27	0.205	-0.02	0.02	5013	-0.97	0.334
Who: Partner	0.03	0.05	5013	0.48	0.628	-0.01	0.04	5013	-0.17	0.866
Who: Friends	-0.01	0.03	5013	-0.38	0.704	0.01	0.02	5013	0.50	0.616
Who: Roommate	0.03	0.04	5013	0.66	0.510	0.05	0.03	5013	1.65	0.100
Who: Patrons	-0.01	0.09	5013	-0.11	0.910	0.06	0.07	5013	0.83	0.409
Who: Colleagues	-0.14	0.04	5013	-4.06	<0.001	0.05	0.03	5013	1.74	0.082



Situation	Positive Affect				Negative Affect				
	Estimate	SE	df	t	Estimate	SE	df	t	p
Who: Boss	-0.06	0.11	5013	-0.51	0.613	0.09	5013	0.67	0.505
Who: Family	0.07	0.07	5013	1.14	0.255	0.05	5013	1.76	0.079
Who: Other	0.15	0.09	5013	1.80	0.071	0.07	5013	-0.77	0.443
Where: Home	0.10	0.03	5014	3.83	<0.001	0.02	5014	-0.31	0.760
Where: School	-0.14	0.03	5014	-5.19	<0.001	0.02	5014	0.81	0.419
Where: Work	-0.08	0.07	5014	-1.13	0.261	0.10	5014	1.97	0.049
Where: Other	0.09	0.04	5014	2.25	0.025	-0.06	5014	-1.85	0.064

Note. N = 419.

**Table 14:**

Between-Person Correlations Between Global Well-Being Judgments and Experiential Measures in Study 2.

	Life Satisfaction	SWLS	Global PA	Global NA
ESM PA	0.44	0.41	0.47	-0.25
ESM NA	-0.33	-0.27	-0.31	0.41
DRM PA	0.34	0.32	0.45	-0.26
DRM NA	-0.27	-0.27	-0.30	0.37

*Note.* SWLS = Satisfaction With Life Scale; PA = Positive Affect; NA = Negative Affect; DRM = Day Reconstruction Method; ESM = Experience Sampling Method. N for correlations with DRM = 461; N for correlations with ESM = 419.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript