9-2020

For Whom, and for What, is Experience Sampling More Accurate Than Retrospective Report?

William D. Ellison
Trinity University, wellison@trinity.edu

Alec C. Trahan
Trinity University, atrahan@trinity.edu

Joseph C. Pinzon
Trinity University, jpinzon@trinity.edu

Megan E. Gillespie
Trinity University, mgillesp@trinity.edu

Logan M. Simmons
Trinity University, lsimmons@trinity.edu

See next page for additional authors

Publication Details

Personality and Individual Differences

Repository Citation

Ellison, W. D., Trahan, A. C., Pinzon, J. C., Gillespie, M. E., Simmons, L. M., & King, K. Y. (2020). For whom, and for what, is experience sampling more accurate than retrospective report? *Personality and Individual Differences, 163*, 110071. doi: 10.1016/j.paid.2020.110071

Authors

William D. Ellison, Alec C. Trahan, Joseph C. Pinzon, Megan E. Gillespie, Logan M. Simmons, and Kendel Y. King

For Whom, and for What, Is Experience Sampling More Accurate than Retrospective Report?

William D. Ellison, Alec C. Trahan, Joseph C. Pinzon, Megan E. Gillespie, Logan M. Simmons, and Kendel Y. King

Trinity University

Author Note

William D. Ellison, Department of Psychology, Trinity University; Alec C. Trahan, Department of Psychology, Trinity University; Joseph C. Pinzon, Department of Psychology, Trinity University; Megan E. Gillespie, Department of Psychology, Trinity University; Logan M. Simmons, Department of Psychology, Trinity University; Kendel Y King, Department of Psychology, Trinity University.

Joseph Pinzon is now at the Department of Psychology, Pepperdine University. Megan Gillespie is now at the Independent Living Research Utilization Program, TiRR Memorial Hermann Research Center.

We thank Kai de Sequera for assistance in developing study materials.

Correspondence concerning this article should be addressed to William D. Ellison, Department of Psychology, Trinity University, One Trinity Place, San Antonio, TX 78212. E-mail: wellison@trinity.edu

Abstract

The experience sampling method (ESM) is often used in research, and promoted for clinical use, with the rationale that it avoids problematic inaccuracies and biases that attend retrospective measures of mental phenomena. Research suggests that averaged scores from ESM data are more accurate than retrospective ratings. However, it is not known how well individuals can remember information about momentary (rather than averaged) mental states, nor how accurately they estimate the dynamic covariation of these states. Individual differences in retrospective accuracy are also poorly understood. In two pre-registered studies, we examined differences between retrospective memory for stress and self-esteem and data gathered via experience sampling and examined whether alexithymia predicted accuracy. Results of both studies revealed substantial discrepancies between retrospective ratings and ESM ratings, especially for momentary states and their covariation. Alexithymia was positively related to recognition of stress means and variability but unrelated to recall of either stress or self-esteem, their variability, or their covariation. These findings suggest that experience sampling may be more useful than self-report when precise information is needed about the timing of mental states and dynamics among them.

For Whom, and For What, Is Experience Sampling More Accurate than Retrospective Report?

The experience sampling method (ESM), also known as ambulatory assessment or ecological momentary assessment, is an increasingly popular way to gather data from individuals in their everyday lives. Scholars have suggested that ESM has the potential to enhance basic psychological research (Miller, 2012), knowledge of psychopathology (Trull & Ebner-Priemer, 2013), and clinical psychological practice (van Os et al., 2017). Indeed, ESM offers a number of potential advantages over retrospective assessment of mental states and behaviors, such as the ability to collect timestamped, contextualized data and the ability to construct person-specific, dynamic portraits of an individual’s functioning (Fisher & Boswell, 2016; Shiffman, Stone, & Hufford, 2008).

Perhaps the most straightforward benefit of ESM is that real-time assessment minimizes the inaccuracy and bias that attend retrospective reports of mental states. Several studies comparing ESM and retrospective measures suggest that memory for past mental states is flawed, a phenomenon known as the “memory-experience gap.” Most commonly, a negative memory bias is found wherein the individual tends to recall more intense or persistent negative experiences than they reported at the time (Ben-Zeev & Young, 2010; Gloster et al., 2008; Kelly, Kertz, Simpson, Bloch, & Pittenger, 2019; Urban et al., 2018; but see Mneimne et al., in press). Some studies suggest that this negativity bias may be a component of a more general extremity bias, in which individuals remember positive experiences as more positive, and negative experiences as more negative, than actually experienced (Ben-Zeev, McHugo, Xie, Dobbins, & Young, 2012; Ben-Zeev, Young, & Madsen, 2009; Neubauer, Scott, Sliwinski, & Smyth, in press; Wenze, Gunthert, & German, 2012). Even so, there is evidence that this bias is larger for negative than for positive experiences (Miron-Shatz, 2009; Ganzach & Yaor, 2019). In general, these studies suggest that averaged experience sampling data provides a more accurate picture of a person's actual thoughts, feelings, and behaviors than relying on retrospective summaries of these experiences.

However, far less research has accumulated regarding the potential benefits of experience sampling for momentary or disaggregated information – that is, momentary states and experiences, rather than averages. Some studies do suggest that retrospective ratings of means and general frequencies are less prone to inaccuracy than estimates of change and instability (Ebner-Priemer, Bohus, & Kuo, 2007, as cited in Ebner-Priemer & Trull, 2009; Stone, Broderick, Shiffman, & Schwartz, 2004). However, these estimates are themselves averages of instability over time and do not directly deal with the timing and sequences of momentary mental experiences. Likewise, some studies suggest that individuals may overestimate how much different mental and behavioral phenomena covary (that is, go together in time) for them (Gloster, Meyer, Witthauer, Lieb, & Mata, 2017; Gloster et al., 2008; Shiffman et al., 1997). However, in general, comparisons between retrospectively assessed and contemporaneously assessed momentary information are few. This is true even though experience sampling is naturally suited to capturing rich momentary data, including not only time-varying psychological phenomena but also their timing, context, and temporal sequence. Indeed, the extent to which retrospective self-report is reliable for these aspects of momentary experience would be of great interest to personality researchers interested in within-person processes (Fleeson & Jayawickreme, 2015; Revelle & Condon, 2015; Vallacher, Read, & Nowak, 2002), as well as to clinicians who work with clients to identify triggers or concomitants of momentary negative thoughts and acute dysphoria.

In addition, there is little research on individual differences that may affect the memory-experience gap. Older individuals may show less distorted recall of aggregated affect ratings (Neubauer, Scott, Sliwinski, & Smyth, in press). Neuroticism, as well as components of neuroticism such as depressivity and anxiousness, may also influence the degree to which individuals overestimate how much negative emotion they rated via experience sampling, in aggregate (Barrett, 1997; Ben-Zeev & Young, 2010; Cutler, Larsen, & Bunce, 1996; Larsen, 1992; Safer, Levine, & Drapalski, 2002; Suls & Howren, 2012). Beyond these specific individual differences, which only pertain to averaged data, little is known about individual differences in the memory-experience gap, and to our knowledge no study has investigated individual differences in memory for momentary (disaggregated) mental states or experiences.

One individual difference that plausibly may influence the utility of ESM in mitigating distorted memory is alexithymia. This is a dimensionally distributed trait (Keefer, Taylor, Parker, & Bagby, 2019; Mattila et al., 2010; Parker, Keefer, Taylor, & Bagby, 2008) that describes the degree to which an individual finds it difficult to identify and describe their own emotional experience, shows limited emotional imagination, and has a cognitive style focused on external events rather than internal states (Taylor & Bagby, 2012). Alexithymia shows correlations with many different deficits in processing emotional information, especially deficits in automatic processing of negative emotions (Donges & Suslow, 2017). Moreover, some recent evidence also suggests that alexithymia’s processing deficits extend to non-affective internal states as well, such as sensation and cognition. Alexithymia relates, for example, to discrepancies between implicit and explicit self-esteem (Dentale, San Martini, De Coro, & Pomponio, 2010) and to poor awareness of bodily sensations (Murphy, Brewer, Hobson, Catmur, & Bird, 2018). It may be the case, therefore, that alexithymia relates to the extent to which individuals process their ongoing internal experiences, both affective and non-affective, and can remember them later. Indeed, many studies suggest that individuals high in this trait have a general deficit in explicit memory for emotional information (Luminet et al., 2006; Meltzer & Nielson, 2010; Suslow, Kersting, & Arolt, 2003; Vermeulen & Luminet, 2009; for a review of this literature, see Apgáua & Jaeger, 2019). Therefore, it is possible that alexithymia may relate to the extent to which ESM improves upon retrospective self-report. However, existing studies have mainly used laboratory-based designs in which self-reported emotions or sensations are captured contemporaneously with objective stimuli or after a short delay (e.g., one hour). This limits the direct relevance of these findings for the utility of ESM in populations both high and low in alexithymia, because experience sampling's mnemonic benefits extend primarily to mental states and events that have occurred at a greater hiatus (e.g., one week). Research also suggests that the timing of alexithymia's effects on memory is important (Muir, Madill, & Brown, 2017). Thus, the specific impact of alexithymia for events that might be captured by experience sampling, instead of self-reported after the amount of time that might elapse in an outpatient clinical context, is not well understood.

The current series of studies aims to fill these gaps in the literature by investigating the memory-experience gap for stress and self-esteem states, in both averaged and disaggregated form, with a one-week gap between ESM ratings and retrospective ones. In both studies, alexithymia is examined as a correlate of accuracy. In the first study, participants completed ratings of stress and self-esteem states 64 times over the course of two days before completing recognition tasks for the averages, changes (timing and magnitude), and dynamic covariances of these states approximately one week later. Study 2 aimed to replicate this design, using recall-based memory tasks instead of recognition.

Study 1

The first aim of study 1 was to investigate the extent of the memory-experience gap for momentary stress and self-esteem states, including both aggregated information (average stress and self-esteem) and disaggregated information (moment-by-moment stress and self-esteem states, and dynamic covariance of stress and self-esteem). The second aim was to investigate the relationship of trait alexithymia to the discrepancy between these variables rated in an experience-sampling protocol and retrospectively assessed values. We hypothesized that alexithymia would exacerbate the memory-experience gap for stress and self-esteem. The study protocol and hypotheses were pre-registered at https://osf.io/ympu6, and study materials and data are also available at the same repository.

Method

Participants and procedure. Participants were recruited randomly, via email, from introductory psychology classes at a small university in the Southwestern U.S. Individual participants attended an initial laboratory visit to complete a demographics questionnaire, receive instructions for completing surveys on their smartphones, and pick a schedule for ESM survey completion. They also completed the Toronto Alexithymia Scale-20 (TAS-20; Bagby, Parker, & Taylor, 1994), a twenty-item self-report measure of alexithymia, at this visit. The TAS-20 is the most frequently used measure of alexithymia and contains three subscales related to difficulty identifying feelings, difficulty describing feelings, and externally-oriented thinking. Items are rated on a 5-point scale. The internal consistency of the TAS-20 in Study 1 was adequate ($\alpha = .81$).

Experience sampling. For the ESM portion of the study, participants completed a survey on 64 occasions using the web browser of their smartphone and Qualtrics online survey software.

A timer application provided auditory prompts for survey completion. Participants were prompted to complete surveys every 15 minutes for eight hours, and this series of 32 surveys was then repeated on a second day. Participants could choose when to start surveys, but they were asked to complete survey series on days when they would be able to do so faithfully. Each survey consisted of two questions, which were rated on a visual analog scale using a sliding response bar and the phone's touchscreen. Self-esteem was measured with the prompt, "Right now, I feel good about myself." Stress was measured with the prompt, "Right now, I feel stressed." Responses were coded using a 0-100 scale with anchors at 0 ("Not at all") and 100 ("Extremely").

After each participant completed these surveys, their 64-occasion dataset was compiled and processed. First, their mean levels of stress and self-esteem were computed. Second, graphical displays (time plots) of their unique stress and self-esteem time series were created. Finally, dynamic factor models were created for each participant, using LISREL, version 8.12 (Jöreskog & Sörbom, 1993), to represent the covariation of stress and self-esteem in contemporaneous ratings and across fifteen-minute intervals. An autoregressive model with a lag of one occasion, in which each variable was regressed on itself at the prior occasion and allowed to correlate with the other variable at the same occasion, was used as the baseline model for each participant. If this model did not show a good fit to the data, modification indices guided the addition of cross-lagged regression parameters between one variable at time $t - 1$ and another variable at time t until satisfactory fit was achieved. Only lag-1 parameters were permitted in these models. Thus, five parameters were possible: the two autoregressive parameters, the two cross-lagged regression parameters, and the contemporaneous correlation between the two variables. Fit decisions were based on cutoffs (Hu & Bentler, 1999) on the Standardized Root

Mean Square Residual (SRMR; value ≤ 0.1), the Root Mean Square Error of Approximation (RMSEA; value ≤ 0.08), the Non-Normed Fit Index (NNFI; value ≥ 0.95), and the Comparative Fit Index (CFI; value ≥ 0.95).

As an additional pre-processing step before these models were created, cubic spline interpolation (Forsythe, Malcolm, & Moler, 1977) with the "spline" function in R software was used to create equal intervals between measurement occasions for each person. This is a requirement of dynamic factor analysis (Molenaar & Rovine, 2011). Cubic spline interpolation fits curves to the observed time series (separately for each variable, day, and individual) and then re-samples from the curves to create time series with equal intervals between observations. In a previous paper using experience sampling, cubic spline interpolation produced model parameters that closely corresponded to those describing the original data (Fisher, Reeves, Lawyer, Medaglia, & Rubel, 2017).

Memory tasks. Participants were asked to return to the lab for a second in-person session approximately one week after completing ESM surveys. They were not told the purpose of this session, except that they would answer some questions about their experience completing mobile surveys. On average, this second session occurred 6.64 days (*SD* = 3.92) after completion of ESM surveys. During the second session, participants were first asked to recall the average level of their stress and self-esteem during the ESM sampling period, on the same 0-100 scale. Then, they were asked to recognize their stress means from among four numbers: their mean and those of three other participants (typically the previous three participants). We used other participants' means in order to provide a realistic set of alternatives for this task. Participants then did the same task for self-esteem. All means were rounded to the nearest tenth of a point. Third, participants were shown four 64-occasion time plots of stress on an 8.5in. x 11in. piece of paper, one of which was theirs and the other three of which were other participants' stress time plots. Participants were reminded of the timing of their surveys and asked to recognize their own stress variation over this time period. This task was also repeated for self-esteem. Finally, participants were shown graphical representations of four dynamic factor models, which represented the longitudinal covariation of stress and self-esteem for themselves (that is, their good-fitting model) and three other models, chosen randomly from among the 30 other possible models (Figure 1). The meaning of the model parameters was explained in detail by a research assistant, who verified each participant's understanding by asking them to explain the parameters back. After participants demonstrated understanding of these models, they were asked to choose their own pattern of dynamic covariation from the alternatives.

Seventy-five individuals (M_{age} = 18.6, age range = 18 to 22) enrolled in the study. Of these participants, 50 (66.7%) reported their gender as female and 25 (33.3%) as male. Five participants (6.7%) reported their race as Black, none as Native American, twelve (16.0%) as Asian, fifty-five (73.3%) as White, and three (4.0%) as multiracial (participants could choose more than one racial category). Twelve participants (19.5%) reported their ethnicity as Hispanic/Latinx. Of the 75 enrolled participants, 13 did not complete both eight-hour survey blocks, and two additional participants did not attend the second lab session. The remaining 60 participants, which was the intended sample size at pre-registration, constituted the sample for Study 1. None of the measured demographic variables (age, gender, race, or ethnicity) related to attrition between enrollment and completion of the second lab session (p-values > .30).

Results and Discussion

Compliance with ESM Protocol and Variability of Responses

Participants were generally compliant with the ESM protocol. They completed an average of 60.13 (*SD* = 5.21), or 94.0%, of the 64 prompted surveys, with a median of 62 surveys completed. In addition, it is important to examine the variability in the data, as the current analyses hinge on memory for both stress and self-esteem means (which vary from individual to individual) and for within-person variability in these variables over time. Specifically, we calculated the between-person standard deviations, the within-person standard deviations, and the root mean square successive differences (RMSSD; e.g., Jahng, Wood, & Trull, 2008) of each variable. The last of these statistics is particularly important for time-series data, as it indexes how variables change over successive surveys (here, over 15 minutes). Stress (SD_{between} = 18.15, SD_{within} = 15.49, RMSSD = 13.90) and self-esteem (SD_{between} = 15.36, SD_{within} = 13.23, RMSSD = 13.76) both showed comparable within-subject and between-subject standard deviations and RMSSD values.[1] This suggests that the memory tasks, in which participants attempted to distinguish their data from others' data, were based on adequate between-person and within-person variability.

Memory Performance and Alexithymia

On average, participants displayed a bias towards recalling more stress (*M* = 41.35, *SD* = 23.07) than they rated in ESM surveys (*M* = 34.20, *SD* = 20.74), *t*(59) = 5.53, *p* < .001, *d* = .33. However, their retrospective recall of self-esteem (*M* = 66.05, *SD* = 18.26) was not significantly biased with respect to ESM ratings (*M* = 65.41, *SD* = 16.63), *t*(59) = .656, *p* = .52, *d* = .04. In recognition tasks, 34 participants (56.7%) correctly recognized their stress mean from among alternatives, which was better than chance, $\chi^2(1) = 30.4$, *p* < .001. Forty participants (66.7%)

---

[1] In response to a comment by an anonymous reviewer, we will note that the ratios of the between-person and within-person variances do not exceed the ratio of 1.5:1, which is cited by Rosopa, Schaffer, & Schroeder (2013) as an indication of heterogeneity of variance.

correctly recognized their self-esteem mean, 40 participants (66.7%) recognized their stress time plot, and 40 (66.7%) recognized their self-esteem time plot (for each of these three tasks, $\chi^2[1] = 53.4$, $p < .001$). In contrast, they performed at chance levels when asked to recognize how self-esteem and stress covaried during the time they were completing ESM surveys: only 14 individuals (23.3%) correctly recognized their own model of covariation from among alternatives, $\chi^2(1) = 0.03$, $p = .86$.

Contrary to hypotheses, participants higher in alexithymia correctly recognized their stress level from among alternatives more often, *OR* = 3.63, *p* = .02, and also recognized their stress time plot from among alternatives more often, *OR* = 3.95, *p* = .02, than those with lower levels of alexithymia (Figure 2). That is, for every unit increase in alexithymia on the 5-point scale with which it was measured, participants' odds of correctly picking their stress level and time plot more than tripled, on average. However, alexithymia did not predict recognition of self-esteem level, *OR* = .798, *p* = .67, or self-esteem time plot, *OR* = 2.12, *p* = .18. Finally, alexithymia did not predict individuals' recognition of their patterns of covariance between stress and self-esteem, *OR* = .856, *p* = .79.

Thus, in general results showed a negative recall bias for stress, but not self-esteem, consistent with the literature on the memory-experience gap for aggregated negative emotional experiences (e.g., Ben-Zeev & Young, 2010). Also consistent with this literature, recognition of self-esteem means was slightly better than recognition of stress means; nevertheless, participants were usually able to recognize both their average stress level and average self-esteem level from among alternatives. Memory for disaggregated stress and self-esteem levels, as measured through recognition tests of stress and self-esteem time plots from among alternatives, was also fair. This suggests that individuals encode and remember a degree of information about their momentary stress and self-esteem states that can sometimes enable them to recognize their own unique fluctuations in these experiences. However, recognition of the dynamic covariation of stress and self-esteem (including how self-esteem and stress states are linked to one another and to themselves across 15-minute intervals) was poor. This finding is consistent with prior studies suggesting systematic discrepancies between the actual within-person covariance of different states and events and individuals' estimation of this covariance (Gloster, et al., 2017; Gloster et al., 2008). The current results suggest that individuals do not perceive and encode these dynamics accurately, giving the experience sampling method (along with within-person analyses) a clear advantage when it comes to uncovering the actual dynamic processes among different mental states, events, and behaviors.

It was particularly unexpected that alexithymia would relate positively to individuals' recognition of their stress means and time plots. One possible explanation for this pattern is that individuals high in alexithymia may have distinctive patterns of stress (for example, different levels or patterns of variability) that may have aided their recognition of these configurations from among alternatives, whereas individuals low in alexithymia may have been unable to distinguish their stress levels and fluctuations from others'. If so, we would expect that this memory advantage for alexithymia would not extend to recall, where there would be no visual cues to help individuals discriminate their experiences from others' in this way. We tested this possibility in a second study, which replicated Study 1 but used only recall-based measures.

Study 2

The results of Study 1 suggested that individuals high in alexithymia more accurately recognized the mean levels and variability in their stress over time. This finding was opposite of what we hypothesized. We reasoned that this surprising effect may have been due to the recognition-based nature of the memory tasks used in this study. If people high in alexithymia had distinctive stress patterns compared to those with more moderate levels of alexithymia, this may have strengthened their ability to recognize these patterns from among alternatives. If so, we reasoned that their relative advantage would not extend to recall-based tasks (Vermeulen & Luminet, 2009). Thus, Study 2 was intended as a replication of Study 1, using only recall-based measures. Like Study 1, the design and hypotheses of Study 2 were preregistered at https://osf.io/ympu6, where study materials and data are also publicly available.

Method

Participants and procedure. Participants were undergraduate students from the same university as in Study 1, and the recruitment method was identical to that of Study 1. Eighty-seven individuals (M_{age} = 18.5, range = 18 to 21 years) enrolled in the study. Of these, 55 (63.2%) were female and 30 (34.5%) were male. Two participants declined to select a gender. Six participants (6.9%) reported their race as Black, one (1.1%) as Native American, eleven (12.6%) as Asian, sixty-six (75.9%) as White, and five (5.7%) as multiracial. Seventeen participants (19.5%) reported their ethnicity as Hispanic/Latinx. Of the 87 participants who enrolled in the study, 12 (13.8%) did not complete the protocol: 10 did not complete ESM surveys, one did surveys but did not attend the second laboratory session, and one participant provided ESM with insufficient variability to be used in the recall procedures. Thus, the final sample consisted of 75 individuals, which was the pre-registered sample size. As in Study 1, all measured demographic variables were unrelated to completion of the study (p-values > .64).

As in Study 1, participants attended an initial lab session, where they completed a demographic questionnaire and the TAS-20. The internal consistency of the TAS-20 in Study 2 was $\alpha = .85$. Thereafter, participants completed 64 surveys measuring their momentary stress and self-esteem before returning to the lab an average of 7.27 days later (*SD* = 5.66). In this second lab session, memory tasks consisted exclusively of recall-based measures. First, participants were asked to recall their average levels of stress and self-esteem during the ESM surveys on a 0-100 scale.

Time plot (recall). Next, participants were given a tablet computer (a Microsoft Surface Pro) on which a 7 in. x 5 in. (17.5 cm x 13.0 cm) rectangular field was displayed. The left side of this field was labeled as a y-axis from 0 to 100. Participants were reminded of the times of the start and end of their first day of surveys and were asked to use the tablet's stylus to trace their memory of stress (or self-esteem) within the field, on the 0-100 scale, from the beginning of their surveys to the end for that day. They then repeated this procedure for the second day of surveys and then again for the second variable (the order in which stress and self-esteem recall was measured was randomized). Their traced time plots were then laid over their actual time plots on an identically sized grid, and the inaccuracy in their recall was quantified as the area between the two lines, in pixels. The two days' pixel counts for stress were then added together to create an overall measure of stress recall inaccuracy; the same was done for self-esteem. These time plots also allowed us to quantify recall bias in average stress and self-esteem in a second way, by comparing the error pixels above their traced plots (underestimation) with the error pixels below (overestimation).

Dynamic covariation (recall). Each individual's bivariate time series was subjected to dynamic factor analysis using the same procedure as Study 1. In the final recall task of the lab visit, a research assistant explained the concept of covariation of stress and self-esteem to participants in basic terms, including the contemporaneous and lagged relationships that appear in these models. This explanation was accompanied by a graphical display (a path diagram without arrows) representing stress and self-esteem at one survey and at the next. Participants were asked to complete the diagram by drawing arrows to represent what they recalled about the covariation between their stress and self-esteem levels during their ESM survey period. Because there were four variables in the model (stress and self-esteem, each at one occasion and the following occasion), five arrows were possible: the contemporaneous, undirected connection between stress and self-esteem; the lagged, autoregressive relationships between stress or self-esteem at one survey and the same variable at the next survey; and the cross-lagged regression relationships, in which stress or self-esteem predicted the other variable at the following survey (Figure 1). After participants had done this, research assistants verified their understanding of the meaning of each parameter by asking the participants to explain them, and any errors in understanding were corrected as needed. After the task was completed, a model consisting of the parameters each participant entered was fit to their actual data. The Standardized Root Mean Square Residual (SRMR) for this model was used to operationalize inaccuracy, as it provides a quantitative index of the degree of model misfit (Maydeu Olivares, 2017). Because the SRMR does not penalize models for overparameterization, participants were instructed to draw the same number of arrows as were needed to achieve adequate fit to their actual data, according to dynamic factor models derived as in Study 1.

Results and Discussion

Compliance with ESM Protocol and Variability of Responses

As in Study 1, compliance with the ESM protocol was generally good. Participants completed an average of 61.56 (SD = 3.28), or 96.2%, of the 64 prompted surveys, with a median of 63 surveys completed. Variability of stress and self-esteem was also adequate, both between individuals and within individuals over a 15-minute interval for both stress (SD_{between} =

18.15, $SD_{\text{within}} = 17.12$, RMSSD = 15.47) and self-esteem ($SD_{\text{between}} = 15.36$, $SD_{\text{within}} = 13.23$, RMSSD = 13.41).

Memory Performance and Alexithymia

Table 1 contains descriptive statistics for the three measures of recall inaccuracy (the difference between participants' recalled means and their means from ESM surveys, the difference between participants' drawings of their stress and self-esteem over time from time plots constructed from ESM data, and the misfit of participants' covariance models when applied to their ESM data), as well as correlations among these indices of inaccuracy. As in Study 1, participants showed a recall bias for stress, recalling higher average stress ($M = 46.95$, $SD = 21.60$) than they reported in experience sampling surveys ($M = 37.82$, $SD = 18.24$), $t(74) = 6.30$, $p < .001$, $d = .46$, 95% CI: 6.25 to 12.03. In contrast to Study 1, participants also recalled significantly lower self-esteem averages ($M = 64.33$, $SD = 18.04$) than they reported during the experience sampling period ($M = 66.36$, $SD = 15.63$), $t(74) = 2.20$, $p = .03$, $d = .12$, 95% CI: -3.87 to -0.19. For comparison, the intraindividual (within-subject) standard deviations for ESM ratings were 17.29 for stress and 13.53 for self-esteem, on average. This means that individuals' ratings were accurate to within 0.52 *SD* for stress and 0.15 *SD* for self-esteem. Thus, consistent with Study 1, individuals were more accurate for self-esteem than for stress.

Figure 3 shows the time-plot drawings for the first eight participants (all of these drawings are available at https://osf.io/ympu6). These drawings also revealed a bias in recalled stress, such that participants overestimated their stress by an average net count (counted as the pixels of overestimation minus the pixels of underestimation) of 137,397 pixels ($SD = 71{,}234$), $t(70) = 4.12$, $p < .001$, $d = 1.93$, 95% CI: 71,224 to 204,651. However, this bias was not statistically significant for the self-esteem time-plot drawings, where participants were generally accurate in the levels of their self-esteem relative to their ESM data (M = 13,685 pixels of underestimation, SD = 220,731), $t(70)$ = .522, p = .60, d = .06, 95% CI: -65,932 to 38,561. Roughly 29% of the area of the stress time-plot drawings consisted of error, whereas this figure was only 18% for self-esteem time-plot drawings.

Participants were generally not accurate in estimating the longitudinal covariance of their stress and self-esteem, with a median SRMR of 0.15. Indeed, fully 75% of the sample had an SRMR that exceeded 0.08, a widely used guideline for good fit (Hu & Bentler, 1999). Comparison of the parameters chosen by participants to parameters appearing in participants' dynamic factor models (Table 2) revealed a systematic bias: participants often underestimated the "inertia" of their stress, rarely recalling that their momentary stress levels tended to relate to their stress levels 15 minutes before. The same was true for self-esteem. In contrast, they overestimated the frequency with which stress would predict self-esteem, and self-esteem would predict stress, over this same time interval. Thus, they tended to overestimate the dynamic, lagged links between these two experiences.

Alexithymia. Participants' level of alexithymia as measured by the TAS-20 did not relate to any of the measures of recall inaccuracy. TAS-20 scores were unrelated to inaccuracy in recall of stress means, r_s = -.06, p = .61, self-esteem means, r_s = -.05, p = .65, stress time plots, r_s = .04, p = .76, self-esteem time plots, r_s = .12, p = .32, and dynamic covariation of stress and self-esteem, r_s = .05, p = .65. Exploratory analyses relating TAS-20 subscales to these outcomes did suggest that the externally oriented thinking subscale related negatively to inaccuracy in recall of stress means, r_s = -.289, p = .012. That is, the higher these scores were, the more accurate individuals were in recalling how stressed they were during the ESM sampling period, on average. Otherwise, none of the TAS-20 subscales were significantly related to any index of recall inaccuracy. In general, these results are consistent with those of Study 1 in suggesting that high levels of alexithymia do not exacerbate the memory-experience gap; in fact, Study 2 provided some limited evidence that a facet of alexithymia might *decrease* the memory-experience gap for stress, but not self-esteem.

General Discussion

The current two studies allow us to quantify the degree to which ESM may be more accurate than retrospective self-report for momentary (disaggregated) psychological data. Participants recognized time plots of their momentary stress and self-esteem data only two-thirds of the time, and roughly 20-30% of the area of participants' drawings of these time plots was "error." Although this degree of inaccuracy does not indicate an absolute advantage for ESM over self-report, it does suggest that in applications where having information about the exact timing of mental states is important, retrospective memory for momentary mental states may not suffice. Timestamped ESM data may be necessary to ensure that that the recovered patterns are accurate. ESM could thus be particularly useful in studies aimed at uncovering within-person processes related to individual differences in personality, as well as in clinical applications where knowledge of both the timing and severity of mental states is essential.

Across two studies, participants displayed a recall bias for stress, exemplified by overestimated stress means in both studies and overestimation of stress in time plot drawings in study 2. This finding is in line with prior research showing a tendency for individuals to recall more negative affect than they reported in real time (Ben-Zeev & Young, 2010; Gloster et al., 2008; Kelly, Kertz, Simpson, Bloch, & Pittenger, 2019; Urban et al., 2018). Somewhat less robust was the recall bias for self-esteem, which was small in size and confined only to study 2. Nonetheless, these findings echo the moderate body of research suggesting particular benefits for aggregated experience sampling data over retrospectively recalled experiences, especially negative ones. Prior research suggests that the size of this bias may depend on the gap between events and recall (the “fading affect bias”; e.g., Gibbons, Lee, & Walker, 2011). Thus, the current study may provide a useful benchmark for the degree to which ESM holds benefits over self-report for aggregated data in some clinically common applications, such as outpatient therapy (where the gap between events and recall often approximates one week).

Memory for how stress and self-esteem covaried over time was poor. Results suggested that participants were merely guessing when asked to recognize these patterns, and that most participants in study 2 did not recall these dynamics with good fidelity, as measured by the median “fit” of the models they chose to describe their own data. Thus, these results imply a consistent advantage for experience sampling over retrospective report for information about dynamic relationships among mental states. This is in line with prior research suggesting that retrospective ratings of means and general frequencies are less prone to inaccuracy than information about change and instability (Ebner-Priemer, Bohus, & Kuo, 2007, as cited in Ebner-Priemer & Trull, 2009; Stone, Broderick, Shiffman, & Schwartz, 2004). This finding also echoes research suggesting disagreements between the actual covariance of events and mental states and individuals’ estimates of this covariation (Gloster, Meyer, Witthauer, Lieb, & Mata, 2017; Gloster et al., 2008; Shiffman et al., 1997). That participants were not only inaccurate in the current study but also biased (in the sense that they recalled more lagged, dynamic relations between stress and self-esteem and underestimated the inertia in these two states) suggests that this inaccuracy may reflect participants’ intuitive or folk-psychological theories of how stress and self-esteem are related. Whatever the source of this inaccuracy, however, the current results strongly suggest that if researchers or clinicians want access to accurate information on the dynamic relationships between mental states, ESM is very likely to outperform self-report.

Our general hypothesis that alexithymia would relate positively to memory inaccuracy for stress and self-esteem was not supported. In fact, alexithymia related positively to recognition of stress levels and stress variation, and the specific alexithymia facet of externally oriented thinking related positively to recall of stress levels. One potential explanation for this finding, as above, is that individuals high in alexithymia may have distinctive levels of stress and patterns of stress reactivity, which may be easier to recognize. This interpretation is supported by the weakening of this effect when recall tasks were employed to measure memory. Another possibility is that high levels of alexithymia may involve a greater attunement to certain aspects of moment-to-moment stress experiences. Alexithymia is related to an attentional focus on external events rather than internal ones (Zimmermann et al., 2005) and a tendency toward an external locus of control (Hexel, 2003; Hungr, Ogrodniczuk, & Sochting, 2016; Verissimo, Taylor, & Bagby, 2000). As a result, it may be that individuals high in alexithymia are more prone to attend to and encode information about *stressors* than those low in alexithymia and may use this information in recognizing their stress level, leading to a relative advantage. This supposition is supported by the advantage in recall of stress that was related specifically to the externally oriented thinking subscale in Study 2. If this is true, it may also help explain why the apparent memory advantage for those high in alexithymia did not extend to self-esteem, which may not track external events as closely. Further research will be needed to confirm these speculative hypotheses, however.

The current study has several limitations that deserve mention. One such limitation is that our sample consisted of university students, which may in some ways be unrepresentative of the general population, as well as the population of individuals involved in outpatient therapy. For example, one potential reason that alexithymia did not negatively impact memory for stress or self-esteem is that the general level of alexithymia in our sample may not have been above the threshold required for these effects to be exhibited. Thus, further research will be needed to explore the extent to which the current results generalize to other groups. A second limitation is that stress and self-esteem were each assessed with only one item in ESM surveys. This raises the possibility that some amount of the discrepancy between individuals' *true* momentary stress and self-esteem and their recalled stress and self-esteem was due to unreliability of these measurements. In future studies, multi-item measures of stress and self-esteem would be preferable to help ensure that apparent memory-experience gaps are not inflated by measurement error. A related limitation concerns the dynamic factor models used. Because of the need not to burden participants excessively, the sampling period was limited to 64 occasions, producing a correspondingly limited degree of statistical power for the parameters appearing in these models. This raises the possibility that participants may have been more accurate in describing the covariance of stress and self-esteem than the current results suggest, especially if they were recalling patterns that are generally true of them but did not characterize the comparatively limited epoch sampled here. For example, if lagged connections between stress and self-esteem were truly present, but weak, our sampling regime may not have recovered them. In addition, because dynamic factor models are contingent on the frequency of measurement, participants may have recalled dynamics between stress and self-esteem that were accurate but which occurred at a speed that was not captured by surveys every 15 minutes. Future research with a greater number of occasions and with a higher sampling frequency will be needed to rule out these competing explanations.

Another potentially fruitful avenue of research would be to combine experience sampling with experimental manipulation of relevant variables, either in a lab setting or in a more ecologically relevant time and place. This would allow for specific states to be induced, which would enable the exact timing of their onset to be recorded and compared with later recall. Future research might also profitably track event-level factors that might impact memory for momentary experiences. The current study considered a person-level predictor of memory for mental states, as have several prior studies (Barrett, 1997; Ben-Zeev & Young, 2010; Cutler, Larsen, & Bunce, 1996; Larsen, 1992; Safer, Levine, & Drapalski, 2002; Suls & Howren, 2012). But disentangling person-level influences on memory accuracy from event-level factors that influence memory is certainly feasible with ESM and would be of considerable interest.

The results of the current series of studies suggest that stress and self-esteem data collected using the experience-sampling method is generally more accurate than retrospectively reported data at a gap of one week. The extent of this advantage for ESM depends on several factors, including whether the retrospective assessment involves recognition or recall, whether aggregated or disaggregated information is measured, and (to a limited extent) the individual's level of alexithymia. These results may have implications for research and clinical practice, especially for applications in which accurate timing is important and in which dynamics among different experiences are a focus of investigation, as this information is particularly difficult for individuals to relay with accuracy. Further research will be needed to clarify the boundaries of these effects.

References

Apgáua, L. T., & Jaeger, A. (2019). Memory for emotional information and alexithymia: A systematic review. *Dementia & Neuropsychologia, 13(1),* 22-30. doi:10.1590/1980-57642018dn13-010003

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38(1)*, 23-32.

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin, 23(10)*, 1100-1110.

Ben-Zeev, D., McHugo, G. J., Xie, H., Dobbins, K., & Young, M. A. (2012). Comparing retrospective reports to real-time/real-place mobile assessments in individuals with schizophrenia and a nonclinical comparison group. *Schizophrenia Bulletin, 38(3)*, 396-404. doi:10.1093/schbul/sbr171

Ben-Zeev, D., & Young, M. A. (2010). Accuracy of hospitalized depressed patients' and healthy controls' retrospective symptom reports: An experience sampling study. *Journal of Nervous and Mental Disease, 198,* 280-285. doi:10.1097/NMD.0b013e318ld614lf

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. *Cognition and Emotion, 23(5)*, 1021-1040. doi:10.1080/02699930802607937

Cutler, S. E., Larsen, R. J., & Bunce, S. C. (1996). Repressive coping style and the experience and recall of emotion: A naturalistic study of daily affect. *Journal of Personality, 64(2)*, 379-405.

Dentale, F., San Martini, P., De Coro, A., & Di Pomponio, I. (2010). Alexithymia increases the discordance between implicit and explicit self-esteem. *Personality and Individual Differences, 49,* 762-767. doi:10.1016/j.paid.2010.06.022

Donges, U.-S., & Suslow, T. (2017). Alexithymia and automatic processing of emotional stimuli: A systematic review. *Reviews in the Neurosciences, 28(3)*, 247–264. doi:10.1515/revneuro-2016-0049

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment, 21,* 463-475.

Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment, 23(4)*, 496-506. doi:10.1177/1073191116638735

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology, 126(8)*, 1044-1056.

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality, 56*, 82-92. doi:10.1016/j.jrp.2014.10.009

Forsythe, G. E., Malcolm, M. A., & Moler, C. B. (1977). *Computer methods for mathematical computations*. Englewood Cliffs, NJ: Prentice Hall.

Ganzach, Y., & Yaor, E. (2019). The retrospective evaluation of positive and negative affect. *Personality and Social Psychology Bulletin, 45(1)*, 93-104. doi:10.1177/0146167218780695

Gibbons, J. A., Lee, S. A., & Walker, W. R. (2011). The fading affect bias begins within 12 hours and persists for 3 months. *Applied Cognitive Psychology, 25,* 663-672. doi:10.1002/acp.1738

Gloster, A. T., Meyer, A. H., Witthauer, C., Lieb, R., & Mata, J. (2017). 'I feel better when…': An analysis of the memory-experience gap for peoples' estimates of the relationship between health behaviors and experiences. *Psychology & Health, 32(9)*, 1152-1166. doi:10.1080/08870446.2017.1327586

Gloster, A. T., Richard, D. C. S., Himle, J., Koch, E., Anson, H., Lokers, L., & Thornton, J. (2008). Accuracy of retrospective memory and covariation estimation in patients with obsessive–compulsive disorder. *Behaviour Research and Therapy, 46,* 642-655. doi:10.1016/j.brat.2008.02.010

Hexel, M. (2003). Alexithymia and attachment style in relation to locus of control. *Personality and Individual Differences, 35,* 1261-1270.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Hungr, C., Ogrodniczuk, J., & Sochting, I. (2016). Alexithymia and locus of control among psychiatric outpatients. *International Journal of Mental Health and Addiction, 14,* 1047-1051. doi:10.1007/s11469-016-9687-x

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods, 13(4)*, 354-375. doi:10.1037/a0014173

Jöreskog, K. G., & Sörbom, D. (1993). LISREL (Version 8.12). Chicago, IL: Scientific Software International.

Keefer, K. V., Taylor, G. J., Parker, J. D. A., & Bagby, R. M. (2019). Taxometric analysis of the Toronto Structured Interview for Alexithymia: Further evidence that alexithymia is a dimensional construct. *Assessment, 26(3)*, 364-374. doi:10.1177/1073191117698220

Kelly, J. M., Kertz, S. J., Simpson, R., Bloch, M. H., & Pittenger, C. (2019). Exploring retrospective biases in obsessive-compulsive disorder: An experience-sampling study. *Journal of Technology in Behavioral Science, 4,* 297-302. doi:10.1007/s41347-018-0078-y

Larsen, R. J. (1992). Neuroticism and selective encoding and recall of symptoms: Evidence from a combined concurrent-retrospective study. *Journal of Personality and Social Psychology, 62(3)*, 480-488.

Luminet, O., Vermeulen, N., Demaret, C., Taylor, G. J., & Bagby, R. M. (2006). Alexithymia and levels of processing: Evidence for an overall deficit in remembering emotion words. *Journal of Research in Personality, 40*, 713-733. doi:10.1016/j.jrp.2005.09.001

Mattila, A. K., Keefer, K. V., Taylor, G. J., Joukamaa, M., Jula, A., Parker, J. D. A., & Bagby, R. M. (2010). Taxometric analysis of alexithymia in a general population sample from Finland. *Personality and Individual Differences, 49,* 216-221.

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika, 82(3)*, 533-558. doi:10.1007/s11336-016-9552-7

Meltzer, M. A., & Nielson, K. A. (2010). Memory for emotionally provocative words in alexithymia: A role for stimulus relevance. *Consciousness and Cognition, 19(4)*, 1062-1068. doi:10.1016/j.concog.2010.05.008

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7(3)*, 221-237. doi:10.1177/1745691612441215

Miron-Shatz, T. (2009). Evaluating multiepisode events: Boundary conditions for the peak-end rule. *Emotion, 9(2)*, 206-213. doi:10.1037/a0015295

Mneimne, M., Furr, R. M., Mendrygal, D., Law, M. K., Arnold, E. M., & Fleeson, W. (in press). Degree of correspondence between retrospective and proximal reports of borderline personality disorder symptoms, symptom triggers, and emotions. *Journal of Personality Disorders*.

Molenaar, P. C. M., & Rovine, M. (2011). *Introduction to multivariate time series analysis.* Pennsylvania State University, University Park, PA.

Muir, K., Madill, A., & Brown, C. (2017). Individual differences in emotional processing and autobiographical memory: Interoceptive awareness and alexithymia in the fading affect bias. *Cognition and Emotion, 31(7)*, 1392-1404. doi:10.1080/02699931.2016.1225005

Murphy, J., Brewer, R., Hobson, H., Catmur, C., & Bird, G. (2018). Is alexithymia characterised by impaired interoception? Further evidence, the importance of control variables, and the problems with the Heartbeat Counting Task. *Biological Psychology, 136,* 189-197. doi:10.1016/j.biopsycho.2018.05.010

Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (in press). How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of Personality and Social Psychology*. doi:10.1037/pspp0000248

Parker, J. D. A., Keefer, K. V., Taylor, G. J., & Bagby, R. M. (2008). Latent structure of the alexithymia construct: A taxometric investigation. *Psychological Assessment, 20(4)*, 385-396. doi:10.1037/a0014262

Revelle, W., & Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality, 56,* 70-81. doi:10.1016/j.jrp.2014.12.006

Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods, 18(3)*, 335-351. doi:10.1037/a0032553

Safer, M. A., Levine, L. J., & Drapalski, A. L. (2002). Distortion in memory for emotions: The contributions of personality and post-event knowledge. *Personality and Social Psychology Bulletin, 28(11)*, 1495-1507. doi:10.1177/014616702237577

Searles, J. S., Helzer, J. E., & Walter, D. E. (2000). Comparison of drinking patterns measured by daily reports and timeline follow back. *Psychology of Addictive Behaviors, 14(3)*, 277-286. doi:10.1037//0893-164X.14.3.277

Shiffman, S., Hufford, M., Hickcox, M., Paty, J. A., Gnys, M., & Kassel, J. D. (1997). Remember that? A comparison of real-time versus retrospective recall of smoking lapses. *Journal of Consulting and Clinical Psychology, 65(2)*, 292-300.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4,* 1-32. doi:10.1146/annurev.clinpsy.3.022806.091415

Stone, A. A., Broderick, J. E., Shiffman, S. S., & Schwartz, J. E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: Absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain, 107,* 61-69.

Stone, A. A., Schwartz, J. E., Broderick, J. E., & Shiffman, S. (2005). Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality and Social Psychology Bulletin, 31(10),* 1340-1346. doi:10.1177/0146167205275615

Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., …Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology, 74(6)*, 1670-1680.

Suls, J., & Howren, M. B. (2012). Understanding the physical-symptom experience: The distinctive contributions of anxiety and depression. *Current Directions in Psychological Science, 21(2)*, 129-134. doi:10.1177/0963721412439298

Suslow, T., Kersting, A., & Arolt, V. (2003). Alexithymia and incidental learning of emotional words. *Psychological Reports, 93,* 1003-1012.

Taylor, G. J., & Bagby, R. M. (2012). The alexithymia personality dimension. In T. A. Widiger (Ed.), *The Oxford handbook of personality disorders* (pp. 648-673). Oxford: Oxford University Press.

Trull, T. J., & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology, 9,* 151-176. doi:10.1146/annurev-clinpsy-050212-185510

Urban, E. J., Charles, C. T., Levine, L. J., & Almeida, D. M. (2018). Depression history and memory bias for specific daily emotions. *PLoS ONE, 13(9)*, e0203574. doi:10.1371/journal.pone.0203574

Vallacher, R. R., Read, S. J., & Nowak, A. (2002). The dynamical perspective in personality and social psychology. *Personality and Social Psychology Review, 6(4)*, 264-273.

van Os, J., Verhagen, S., Marsman, A., Peeters, F., Bak, M., Marcelis, M., …Delespaul, P. (2017). The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and Anxiety, 34,* 481-493. doi:10.1002/da.22647

Verissimo, R., Taylor, G. J., & Bagby, R. M. (2000). Relationship between alexithymia and locus of control. *New Trends in Experimental and Clinical Psychiatry, 16*, 11-16.

Vermeulen, N., & Luminet, O. (2009). Alexithymia factors and memory performances for neutral and emotional words. *Personality and Individual Differences, 47*, 305-309. doi:10.1016/j.paid.2009.03.018

Wenze, S. J., Gunthert, K. C., & German, R. E. (2012). Biases in affective forecasting and recall in individuals with depression and anxiety symptoms. *Personality and Social Psychology Bulletin, 38(7)*, 895-906. doi:10.1177/0146167212447242

Zimmermann, G., Rossier, J., de Stadelhofen, F. M., & Gaillard, F. (2005). Alexithymia assessment and relations with dimensions of personality. *European Journal of Psychological Assessment, 21(1)*, 23-33. doi:10.1027/1015-5759.21.1.23

Table 1

Spearman's Rank Correlations among Measures of Recall Inaccuracy in Study 2

Measure	1.	2.	3.	4.	5.	*Mdn*	*IQR*
1. Stress mean	-					8.32	3.87 - 13.63
2. Self-esteem mean	.205	-				4.49	2.52 - 7.92
3. Stress time plot	.446***	.099	-			330,389	205,609 – 462,490
4. Self-esteem time plot	.402**	.308**	.467***	-		234,741	141,238 – 346,571
5. Dynamic covariation	-.025	.002	.021	-.043	-	.1507	.088 - .234

Note. Inaccuracy for means represents the absolute difference between the means of ESM data and the recalled means, on a 0-100 scale. Inaccuracy for time plots represents the number of pixels between the time plots of ESM data and plots drawn by participants, out of a maximum of 2,124,690. Inaccuracy for dynamic covariation represents the standardized root mean square residual (SRMR) when participant-generated covariance models were fit to their ESM data.

Table 2

Frequency of Dynamic Covariance Parameters Recalled by Participants and Describing Participants' ESM Data

	Recalled		In model	
Parameter	*n*	%	*n*	%
1. Autoregression: stress predicts later stress	42	56	61	82
2. Autoregression: SE predicts later SE	31	41	53	72
3. Cross-lagged regression: Stress predicts later SE	44	59	16	22
4. Cross-lagged regression: SE predicts later stress	22	29	6	8
5. Correlation of stress and SE contemporaneously	54	72	57	77

Note. $N = 75$ for recall task and $N = 74$ for models (one participant's data could not be modeled).

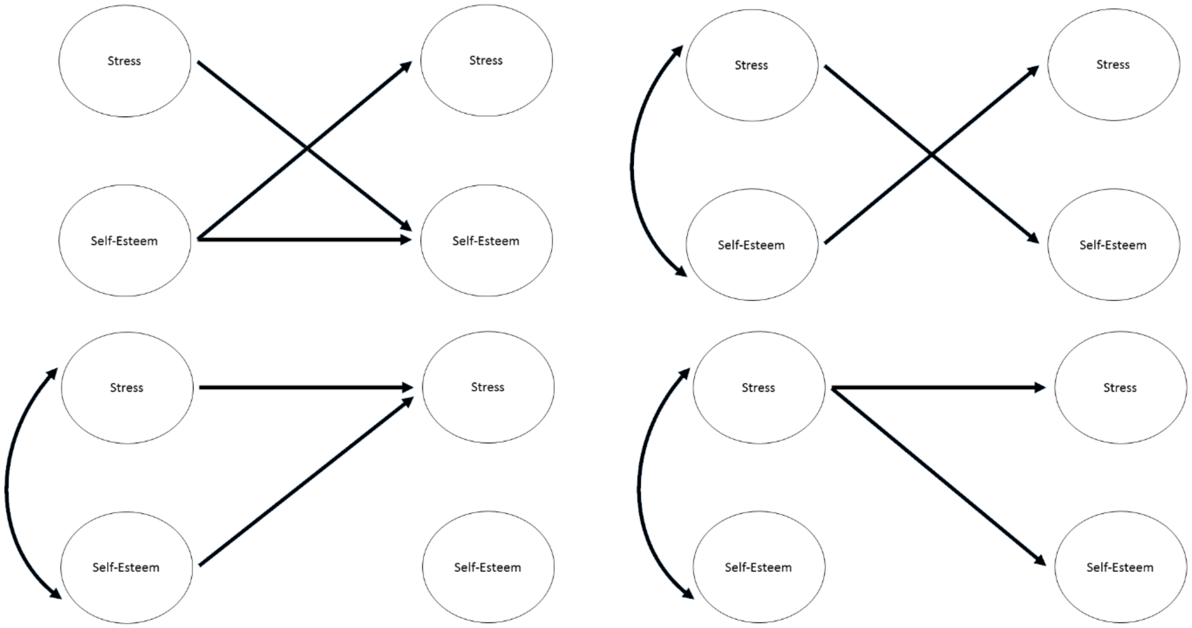

Figure 1. Four models of dynamic covariance among stress and self-esteem. Straight lines represent regression parameters in which stress and/or self-esteem predict these states 15 minutes later, and curved lines represent correlation parameters in which stress and self-esteem covary in contemporaneous surveys.

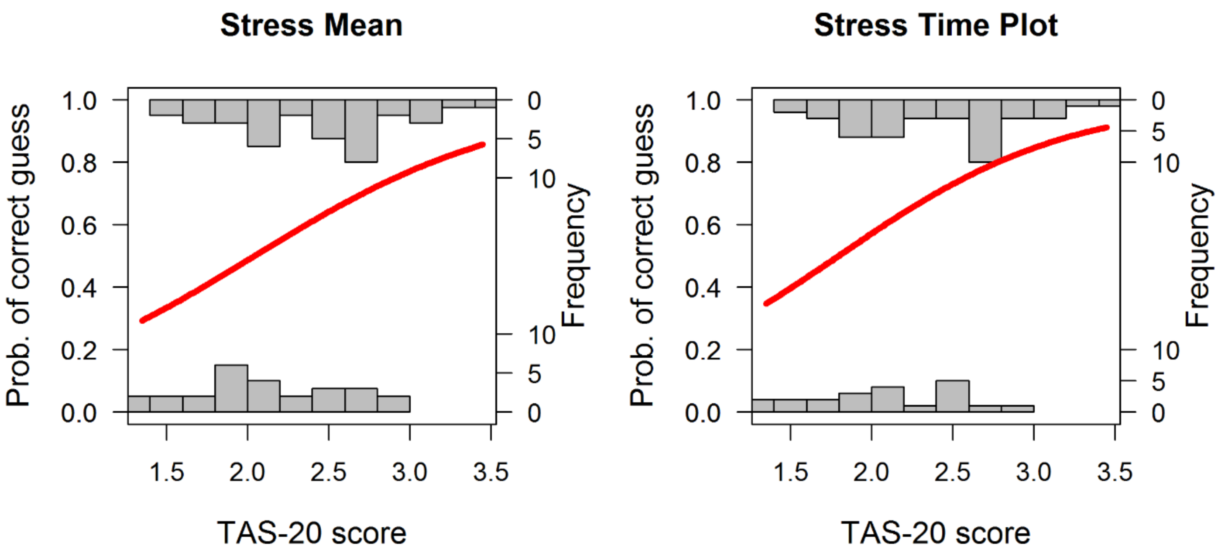

Figure 2. Histograms of TAS-20 scores (scale means) for individuals who correctly recognized (top) and incorrectly recognized (bottom) their mean stress level and stress time plot from among alternatives in Study 1. The red lines represent logistic regression curves predicting correct guesses from alexithymia.

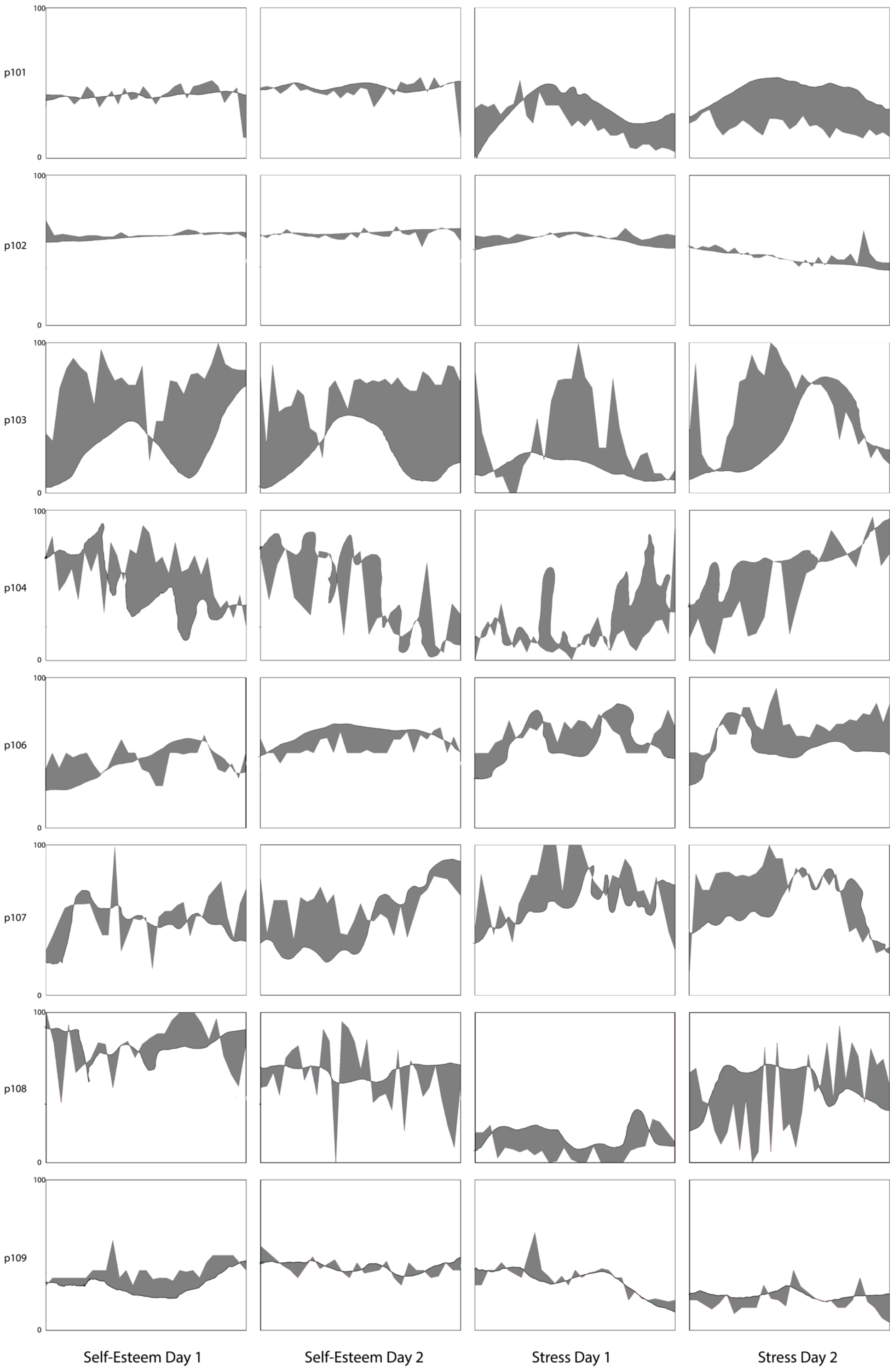

Figure 3. Time plot drawings for the first eight participants, showing recall of momentary self-esteem and stress throughout successive 8-hour ESM survey bursts against the actual values. The gray area represents the discrepancy in recall. ID numbers at left show the participants whose data are represented. Plots for the remaining participants are available at https://osf.io/ympu6.