Global Reports of Well-Being Overestimate Aggregated Daily States of Well-Being

David B. Newman[1], **Norbert Schwarz**[1], **Arthur A. Stone**[1]

[1]University of Southern California

Abstract

Researchers can characterize people's well-being by asking them to provide global evaluations of large parts of their life at one time or by obtaining repeated assessments during their daily lives. Global evaluations are reconstructions that are influenced by peak, recent, and frequently occurring states, whereas daily reports reflect naturally occurring variations in daily life. The present research compared the averages of individual global evaluations and corresponding aggregated daily states from an ordinary two-week period and used a range of well-being measures (life satisfaction, meaning in life, and affect) and related constructs (searching for meaning in life and nostalgia). Across all measures, global reports were significantly higher than aggregated daily states. That is, life is considered more satisfying, more meaningful, and is characterized to a greater extent by more intense positive and negative emotions when reflecting on life in general than when reflecting on daily life in real time.

Keywords

well-being; daily diary; ecological validity; global evaluations; emotion

What can global evaluations and reports of daily experiences tell us about how someone's life is going? Global evaluations require people to bring to mind relevant aspects of their lives and privilege memorable experiences, such as a wedding or an exotic vacation. Single assessments or recollections of life-in-general necessarily omit ordinary and mundane experiences, such as sitting in front of the computer. Assessments in daily life likely capture the quotidian ebbs and flows of someone's life but may miss some of the peak memorable experiences. Understanding the distinction between global evaluations of one's life and contextualized reports of states has been an important topic in social and personality psychology. Well-being can be assessed in similar manners.

These distinct methods of assessing well-being have various strengths and weaknesses. The primary goal of the present research was to compare the averages of specific global reports with their corresponding aggregated daily states of well-being. A secondary goal was to examine the effect of the order of completion of the daily and global reports on the discrepancy between the two. Doing so provides insight into the cognitive processes

Correspondence concerning this article should be addressed to David B. Newman, USC Dornsife Center for Self-Report Science, 635 Downey Way, VPD 405, Los Angeles, CA 90089-3332. Contact: davidnew@usc.edu.

involved in each form of judgment and yields provocative implications for positive psychology, namely that these different methods of measuring well-being do not capture the same construct.

Processes Involved in Global Reports and Aggregated States

Global reports of well-being (e.g., "All things considered, how satisfied are you with your life as a whole these days?"; World Values Survey) allow people to consider longer periods of time that might include peak experiences or major life events, such as marriage or career accomplishments. Clearly people are not able to replay their entire life as they make these judgments. Rather, people are influenced by chronically accessible information and temporarily accessible information (Schimmack, Diener, & Oishi, 2002; Schwarz & Strack, 1999). Chronically accessible information refers to any type of information that is brought to mind repeatedly over time. Temporarily accessible information can take the form of one's present mood (Schwarz & Clore, 1983), thoughts about your current romantic situation (Schwarz, Strack, & Mai, 1991), or thoughts about the current state of the country (Deaton & Stone, 2016). Chronically as well as temporarily accessible sources of information can have profound effects on subsequent judgments of global well-being reports.

This means that global evaluations are based on inputs that are retrieved from memory at the time of judgment. The longer the reflection period, the more extreme instances people will be able to find. This essentially produces a peak effect. Additionally, the currently accessible information, such as how one's life if going right now, will influence this judgment. The net effect results in a judgment that is highly influenced by memorable peak experiences that are likely come to mind, tempered by the current situation, which is rarely as extreme as the peak experience.

Studies that have asked people to recall a recent past experience support this process model. For example, in one study, participants completed momentary reports of their satisfaction and positive and negative affect at randomly selected times during a spring break vacation (Wirtz, Kruger, Napa Scollon, & Diener, 2003). Retrospective reports of each specific measure of well-being were higher than the corresponding momentary averages. Similarly, global reports of how one typically feels while driving luxury cars are typically higher and not strongly related to episodic reports of a recent driving experience (Schwarz & Xu, 2011). In other studies, participants have been asked to recall their well-being, pain, fatigue or other health relevant variables from the preceding few days or weeks (Stone et al., 1998; Stone, Broderick, & Schwartz, 2010). When the momentary average is compared with the recall, the recall is typically higher than the momentary average (Broderick et al., 2008; Stone et al., 2010; Stone, Broderick, Shiffman, & Schwartz, 2004). For example, participants completed momentary reports of pain and fatigue at randomly selected times during the day, and they also completed end-of-day reports of their pain and fatigue. End-of-day reports were significantly higher than the average of momentary reports (Stone et al., 2010).

In addition to using global reports to characterize individuals, researchers can measure individual differences by aggregating daily or momentary states. This method has been advocated among researchers in various areas, including social interactions, romantic

intimacy, personality, and health behaviors (e.g., Affleck, Zautra, Tennen, & Armeli, 1999; Bolger, Davis, & Rafaeli, 2003; Fleeson, 2001; Tennen, Suls, & Affleck, 1991) and could theoretically be applied to individual differences in well-being. Doing so would eliminate or drastically reduce any recall bias inherent in global reports. It would also incorporate the situation more appropriately than global evaluations. Assuming the sample of time points constitutes a fairly representative sample for this period of the person's life, an aggregation of these states may provide a characterization of the individual that closely aligns with individual differences in actual daily experiences.

Although perhaps more "accurate" in describing the attributes of an individual, momentary or daily reports likely capture more mundane and less memorable experiences than global reports. Hence, they may not represent the way people think about themselves and they may not predict future intentions as well as global reports (e.g., Conner & Feldman Barrett, 2012; Wirtz et al., 2003).

The Present Research

Previous research in this area has addressed different questions than the questions posed by the current study. Much of the prior work has focused on the correlation of aggregated daily or momentary states and global evaluations or recalls. Although momentary states measured throughout the course of the day are different from end-of-day reports of states, researchers have attempted to characterize individuals based on the aggregation of both types of measurement. For example, correlations between global reports of affect and daily averages of affect ranged from .59 to .65 (Newman, Nezlek, & Thrash, 2018). In other health domains, the correlation between the recall and momentary average of pain and fatigue has ranged from .81 to .90 (Stone et al., 2010).

The present research extends prior findings in several important ways. First, whereas previous research has compared mean levels of aggregated daily or momentary states with recall of that experience, the present study compares aggregated daily states with global ratings. Second, the present research expands the range of measures that has typically been considered among studies that measure well-being. Well-being can be reasonably conceptualized as containing three aspects: evaluative (e.g., how satisfied people are with their lives), experiential (e.g., how people feel when living their lives), and eudaimonic (e.g., how meaningful or purposeful their lives seem; see Kahneman, 1999; Schwarz & Strack, 1999; Steptoe, Deaton, & Stone, 2015, for a discussion). In addition to measuring evaluative and experiential well-being (measures that have dominated the well-being literature; see Diener, Lucas, & Oishi, 2018; Diener, Suh, Lucas, & Smith, 1999), we measured a sense of meaning and purpose in life. Moreover, we sought to examine two relevant constructs that have less normal distributions in daily life, nostalgia and searching for meaning in life. The measurement of these additional constructs allowed us to generalize our findings more broadly. Third, the present study counterbalanced the order in which participants completed the global reports and the daily reports, which allows for the possible examination of any order effect and provides insight into how global judgments are formed. Given that people are influenced by peak experiences, recent episodes, and frequently occurring events and states, global reports should be more highly correlated with peak, recent, and average states

among participants who complete the daily diary reports before the global evaluations. When daily states are brought to mind through repeated administration, they should influence a subsequent judgment (e.g., Schwarz et al., 1991). Many factors may influence global evaluations of well-being, and the present study builds on this literature by considering how daily states of well-being when brought to mind can influence global well-being evaluations.

Method

Participants and Procedure

We conducted an initial study and replicated the findings with a larger sample. We recruited as many participants as possible within the university participant pool constraints. After data collection, we conducted power analyses and used the observed sample effect sizes as estimates of the population effect sizes (O'Keefe, 2007). Because accurate effect sizes could not be calculated from multilevel models, we relied on effect size estimates from general linear model estimates which likely underestimate true effect sizes. Using the statistical program GPower (Faul, Erdfelder, Buchner, & Lang, 2009), the power analyses indicated that the main effects testing the primary analyses were highly powered (.99–1.00) for all effect sizes, which ranged from $d = .29$ to $d = 1.00$. A sensitivity analyses showed that we had .80 power to detect small effects ($d = .18$). Similarly, the interaction effects were highly powered (.99–1.00) to detect observed small effects ($d = .20$) and all effects larger than $d = .20$. A sensitivity analyses indicated that we had .80 power to detect small effect sizes ($d = .15$). The correlation comparisons were reasonably powered (.80) to detect medium effect sizes ($q = .36$). In short, our samples were adequately powered to detect the effects of interest. The two studies had very similar procedures and measures across two semesters. Because the results were very similar across semesters and for the sake of brevity, we describe the method and results of both studies together.

Participants were undergraduate students who received course research credit. They were instructed to think about their life as a whole as they completed a questionnaire at one time. They also received daily questionnaires for two weeks and were instructed to think about their day as they completed the daily questionnaires. Daily questionnaires were emailed to students at 9:00pm each evening. Reminder emails were distributed the following morning at 7:00am to those who did not complete the questionnaire in the evening. Responses were accepted until 10:00am. Daily questionnaires were eliminated if they were completed after 10:00am, if they were duplicate entries, if the participant failed to correctly answer an instructed response item as recommended by Meade and Craig (2012), or if the participant completed less than five valid entries[1]. Of the initial 3,153 daily questionnaires, we used 2,984 (5.36% eliminated) daily questionnaires in our final analyses, which included 244 participants ($M_{age} = 20.21$; $SD = 2.09$; 74.2% female). On average, participants completed 12.23 daily reports ($SD = 2.12$; minimum = 5, median = 13). The participants were also randomly assigned to either complete the global evaluations before or after the two-week

---

[1]Because daily questionnaires in the second study were longer than the daily questionnaires in the first study, daily questionnaires from the second study completed in less than two minutes were also eliminated from final analyses.

diary portion of the study. The global evaluations questionnaire was administered two days before or two days after the two-week daily diary period.

Measures

Global ratings of satisfaction with life were measured with the 5-item Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) with responses recorded on a 7-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*). Global ratings of presence and search for meaning in life were assessed with an adapted version of the 10-item Meaning in Life Questionnaire (Steger, Frazier, Oishi, & Kaler, 2006). The first item was changed from "I understand my life's meaning" to "My life is full of meaning," and the second item was changed from "I am looking for something that makes my life feel meaningful" to "I am searching for something that makes my life feel meaningful." These items were altered so that the corresponding reworded items administered in the daily questionnaires would make more sense. Responses were recorded on a 7-point scale (1 = *Absolutely untrue*, 7 = *Absolutely true*). Global ratings of positive and negative affect were measured with items that have been used in circumplex models that distinguish valence and arousal (e.g., Brandstätter, 2007; Feldman-Barrett & Russell, 1998; Nezlek, 2005). Positive activated (PA) affect items were enthusiastic, delighted, happy, glad, and excited; positive deactivated (PD) items were calm, peaceful, relaxed, contented, and at ease. Negative activated (NA) affect items were stressed, angry, annoyed, tense, and nervous; negative deactivated (ND) affect items were depressed, disappointed, miserable, gloomy, and sad. Responses were recorded on a 7-point scale (1 = *do not feel this way at all*, 4 = *feel this way moderately*, 7 = *feel this way very strongly*). Global ratings of nostalgia were assessed with the 4-item PINE scale (e.g., "How nostalgic do you feel?", "To what extent do you feel sentimental for the past?"; Newman, Sachs, Stone, & Schwarz, in press). Responses were recorded on a 7-point scale (1 = *Not at all*, 7 = *Very much*).

Daily well-being measures were assessed by rewording several of the corresponding global items as is common practice in diary studies (Nezlek, 2012, pg. 32–33). Participants answered questions about their daily affect using the same items that were used in the global questionnaire, and responses were again recorded on a 7-point scale that reflected the daily nature of the questions (1 = *did not feel this way at all*, 4 = *felt this way moderately*, 7 = *felt this way very strongly*). Daily nostalgia was measured with the PINE scale, but questions were reworded to be appropriate for daily reports, e.g., "How nostalgic did you feel today?" Responses were recorded on a 7-point scale (1 = *Not at all*, 7 = *Very much*). Daily satisfaction with life was measured with two items that were adapted from the Satisfaction with Life Questionnaire ("I was satisfied with my life today" and "The conditions of my life today were excellent"). Responses were recorded on a 7-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*). Presence and search for meaning in life were also measured with two items each that were adapted from the Meaning in Life Questionnaire (presence: "My life today was full of meaning" and "My life had a clear sense of purpose today"; search: "Today, I was searching for something that makes my life feel meaningful" and "I was looking to find purpose in my life today") (Steger et al., 2006). Responses were also recorded on a 7-point scale (1 = *Absolutely untrue*, 7 = *Absolutely true*). Response scales for

the daily measures were the same as the response scales for the corresponding global scores so that the averages could be compared.

Because the daily questions about meaning in life and satisfaction with life contained the word "life," we tested for any ambiguities in interpretation by randomly assigning half of the participants in the second sample to answer the same questions reworded with the word "day" instead of "life." For example, the first item measuring presence of meaning in life was "My life (day) today was full of meaning." The wording of these items had no substantive effect on the averages, so the analyses presented below reflect the aggregated scores across conditions.

It is also important to note that given the added burden of repeated daily questionnaires on participants, we could not administer all of the daily items for satisfaction and meaning that we administered for the global reports. To facilitate an appropriate comparison of the daily reports and the global reports, we selected the two items from the global reports that mirrored the daily report questions as reasonable indicators of the construct. Analyses that follow reflect the two-item global reports of satisfaction with life, presence of meaning in life, and search for meaning in life.

Results

Preliminary Analyses

Descriptive statistics and reliability estimates from Cronbach's alpha for all global reports are reported in Table 1. To calculate descriptive statistics for the daily reports, we used multilevel modeling to account for between- and within-person variation and used the program HLM 7.01 (Raudenbush, Bryk, & Congdon, 2011). Reliabilities of the daily measures were calculated following the guidelines by Nezlek (2017). We first created an item level file that contained the scores of the individual items. Items were nested within days, and days were nested within persons. The item variable for each measure was entered as the outcome measure in separate unconditional models. The reliability of the intercept provides a ratio of true variance over total variance, a common definition of reliability (Nezlek, 2017). Validity refers to the correlation between a global report and the corresponding daily aggregate. The decrease in level 2 variation from the model that included the global report was divided by the total level 2 variation from the unconditional model. The square root of this is analogous to a Pearson's correlation. All daily measures had reasonably high reliabilities (See Table 1).

Next, we ran two-level (days nested within people) unconditional models to examine the amount of within- and between-person variation (See Table 1). These analyses showed that a considerable amount of variation occurred at both levels. Next, the corresponding global measure was entered at level 2 to examine the correlation between the global score and the average of the corresponding daily measure. The correlations between the global measure and corresponding daily averages were all reasonably high with the exception of searching for meaning in life.

Primary Analyses

Critical to the main purpose of the study, we compared the means of the global reports with the corresponding aggregated means of the daily reports (See Table 2). To do so, we subtracted each daily score from that person's corresponding global evaluation score. In essence, daily scores were centered around each individual's global evaluation score. We then built unconditional models in which this new difference score variable was entered as the outcome measure. The intercept was allowed to vary randomly, and no predictors were entered into the model. Error terms with random effect p-values greater than .15 were trimmed from the models as recommended by Nezlek (2012, pp. 65–68). The intercept of this model provides an estimated difference between the global reports and the aggregated daily scores while maintaining the nested data structure. There were significant main effects for all variables such that global reports were higher than daily averages.

Although all global measures were higher than aggregated daily states, some differences were stronger than others. The global-state difference was smallest (.30) for satisfaction with life, followed by PD (.41) and the presence of meaning in life (.55). To compare these differences statistically, we "stacked the data" by creating an outcome variable that alternated between two difference scores (see Bauer, Preacher, & Gil, 2006, for a discussion of a similar procedure). A dummy coded variable representing one of the difference scores was entered uncentered at level 1 and the level 1 intercept was kept. The coefficient of the dummy-coded variable represents the difference between the respective difference scores. These models indicated that the satisfaction with life global-state difference was not significantly different from the PD global-state difference, $\gamma_{10} = -.12$, $t = 1.59$, $p = .112$, pseudo-$R^2 = .19$[2], but it was significantly different from the presence of meaning in life global-state difference, $\gamma_{10} = -.25$, $t = 3.23$, $p = .001$, pseudo-$R^2 = .20$, and from all other measures (all $\gamma_{10} < -.30$, ps $< .003$, pseudo-R^2s $> .19$). These results show that the discrepancy between global reports of satisfaction with life and aggregated daily states of satisfaction is not as large as the other respective well-being discrepancies.

Next, we examined the interaction of order (global reports completed before vs. after) by measure (global vs. daily), which was significant for most measures (see Table 2). The difference between global reports and aggregated daily states was larger among people who completed the global reports before the daily diaries. The one exception to this pattern was satisfaction with life. The discrepancy between global satisfaction and aggregated daily satisfaction was larger among those who completed the global reports after the daily diaries. This interaction was most pronounced in the analyses involving the negative constructs (negative activated and deactivated affect, searching for meaning in life, and nostalgia).

In addition to examining differences in means, we compared the correlations between 1) peaks (maximum values), recent states (last three daily reports), and averages of daily states and 2) global reports across conditions. As expected, many of the correlations were stronger among those who completed the global reports after the daily diary reports (see Table 3). These differences were most pronounced in the correlations involving peak states and were

---

[2]An effect size estimate, pseudo-R^2, was calculated by taking the percent the within-person variance from the null model was reduced when the dummy-coded variable was entered as a predictor at Level 1.

more evident in measures of satisfaction with life, meaning in life, and nostalgia than they were in affective states. Considering the results in Table 3 altogether, it is important to stress the overall pattern of effects instead of attempting to categorize single specific effects based on arbitrary p-value thresholds (Amrhein, Greenland, & McShane, 2019). It is worth noting that many of these effects were small in magnitude though. These findings reiterate the difference between the aggregation of daily states, which are influenced by the present day, and global evaluations, which can be influenced by peak states, recent states, and frequently occurring states in daily life.

Discussion

The present study investigated two different methods of depicting how people's lives are going. Global evaluations of satisfaction with life, affect, presence of and search for meaning in life, and nostalgia were consistently higher than their corresponding daily aggregates. This confirms that people do not accurately replay their entire lives to create an average report of how their lives have been (Bradburn, Rips, & Shevell, 1987). Rather, they appear to reflect on key aspects of their lives, such as peak experiences and recent memories, and they can be influenced by various additional factors, such as personality traits and standards of comparison. Because daily life is more mundane and less intense than the most memorable moments we experience over extended periods of time, the aggregation of daily reports suggests a more moderate life than one might expect on the basis of global reports.

The discrepancy between global evaluations and daily states is consistent with and builds upon previous research which has compared the "experiencing self" to the "remembering self" (Kahneman & Riis, 2005) but also extends the research in important ways. Much of the previous research has compared recollections of specific experiences, such as spring break vacations (Wirtz et al., 2003), menstrual cycles (McFarland, Ross, & DeCourville, 1989), or EMA reporting periods (Stone et al., 1998), with aggregated states during the experience. Other studies have compared global reports of specific aspects of their lives (e.g., math anxiety) with online experiences of the specific event (e.g., taking a math test; Goetz, Bieg, Lüdtke, Pekrun, & Hall, 2013). The general pattern that has emerged in these studies is that as psychological distance increases, report levels tend to increase, presumably because the level of abstraction has also increased. Additionally, recall tends to converge with lay beliefs about how their experiences typically are or should be.

The present study has extended the research in this area by comparing global evaluations with aggregated states, as opposed to recalls of specific periods. It has also increased the range of well-being measures to include meaning in life. This has yielded several implications concerning the ways people think about their lives, the underlying processes involved in these judgments, and the methods used to characterize the well-being of individuals.

Implications for Thinking about Your Life

One of the key implications of these studies concerns the distinction of the interpretations of the results between positively- and negatively-valenced well-being measures. When reflecting on life in general, life seems more meaningful, more satisfying, and more

intensely full of positive emotions. This may lead some to conclude that when evaluating your life, it is best to think of the big picture. However, doing so would also lead to the conclusion that life is more stressful, sad, and depressing than it is in daily life.

It is also worth noting that the discrepancy between global reports and aggregated daily states was considerably larger for negative activated and deactivated affect, searching for meaning in life, and nostalgia than the positively-valenced variables. This could be due to the non-normal distributions of each variable in daily life. Moreover, the correlation between aggregated daily states and global reports of searching for meaning in life was considerably lower than other variables. This suggests that people think about different aspects of their lives when they search for meaning in their lives in general as opposed to searching for meaning in their daily lives, consistent with the results from Newman et al. (2018).

The results also have important implications for understanding how people think about what brings them meaning and purpose in life. For example, in a thorough review of published reports of global reports of meaning in life, Heintzelman and King (2014) showed that the average is higher than the midpoint of the scale and concluded that life is actually pretty meaningful. The present study qualifies this conclusion by noting that life is pretty meaningful when people consider their life as a whole, but it is less meaningful when people consider their daily experiences. This discrepancy between global meaning in life and daily meaning in life is consistent with some recent experimental findings that have shown that life is more meaningful as psychological distance is increased, for example by imagining oneself in a distant location or by thinking about oneself in the future or the past (Waytz, Hershfield, & Tamir, 2015). Our results suggest that meaning can be found at higher levels when thinking about one's life more broadly, which allows for the recollection of rare experiences that may imbue one's life overall with meaning but may not provide much meaning from day-to-day during the mundane activities of daily life.

Our findings also contribute to the growing body of research that compares specific eudaimonic measures of well-being (e.g., meaning and purpose in life) with evaluative and experiential measures of well-being (e.g., Henderson, Knight, & Richardson, 2013; Kopperud & Vittersø, 2008; Newman, Schwarz, Graham, & Stone, 2019; Nezlek, Newman, & Thrash, 2017; Tov & Lee, 2016). These comparisons have important implications for how people think about their lives. The discrepancy between global and daily meaning was more similar to the affect discrepancy than the satisfaction discrepancy. Whereas previous research has shown that daily affect is more highly related to daily satisfaction than daily meaning in life (Tov & Lee, 2016), our findings show that the affect discrepancy is more similar to the meaning in life discrepancy than the satisfaction discrepancy. However, the meaning in life discrepancy needs to be interpreted differently. There is no gold standard for meaning in life. That is, how meaningful someone's life is at the present (e.g., while writing a dissertation) may depend on future life circumstances (e.g., whether that person becomes a successful professor).

Implications for Underlying Processes

By counterbalancing the order in which participants completed the global and daily reports, we were able to glean some useful insights into the underlying processes. Although the

results from the interactions and correlation comparisons were not as robust as the main effects, the overall pattern of findings suggests that accessible information can influence a subsequent judgment (Bless & Schwarz, 2010). For instance, the discrepancy between global and daily aggregated states was strongest among those who completed the global reports before the daily diary, and this discrepancy was most pronounced for negative activated and deactivated affect, searching for meaning in life and nostalgia (satisfaction with life actually followed the opposite pattern). This pattern of results could be explained from a consistency motivation account (Ross, 1989). Among those who completed the diary reports first, those particular days provide a more accessible input than the distant days, which would only be accessible if they were extreme. The daily reporting attenuates the advantage of memorable episodes at the expense of more recent episodes.

Similarly, the stronger correlations between peak, recent, and average daily states and global reports among those who completed the daily reports before the global reports attest to this account. The findings are consistent with the notion that global evaluations involve the reflection of peak experiences, whereas daily reports do not (Schwarz, 2012).

Methodological Implications

Our results have valuable methodological implications as well, particularly for researchers who wish to measure individual differences in well-being with methods often used by personality psychologists. For example, is it more appropriate to measure personality traits by asking people to reflect on their life at one time or should researchers measure momentary or daily states of personality and examine the distribution of these states? The latter method, referred to as Whole Trait theory, has been advocated recently as a method of measuring individual differences in personality traits while still capturing within-person variation (Fleeson & Jayawickreme, 2015). Although beneficial in many regards, aggregating daily or momentary states may miss peak experiences, which happen rarely but nevertheless influence people's perceptions of themselves. Aggregating daily states may not capture individual differences in important life events that do not occur on a daily basis.

Relatedly, the correlation between aggregated daily or momentary states and the global report of that particular variable has been a common metric of scale validation in personality psychology. Correlations in the range of .60 to .70 have typically been considered a reasonable standard for convergent validity. The present results showed that although the correlations between aggregated daily states of well-being and global reports of well-being were reasonably high, these correlations were influenced by the order in which these measures were completed. When global reports are completed after repeated daily reports, the correlation between the two are stronger. Thus, the order in which these measures are completed should be considered.

Limitations and Future Directions

One limitation of the study is that the two-week daily diary period may not be long enough to capture a truly random and representative sample of days of the year. The goal of the study was to capture typical days of the year. If we had included atypical days, such as the final exam period or a spring break vacation, average levels of well-being may have differed,

which would alter the discrepancy between daily aggregates and global reports. Comparing global reports with daily averages over longer periods of time that include unique experiences remains a fruitful avenue for future research.

Another limitation of the study was that participants' affective reports were assessed at the end of the day rather than in the moment. This daily recall introduces a certain amount of bias (Shiffman, Stone, & Hufford, 2008). In future studies, aggregated affective states could be measured with the use of EMA techniques.

Finally, participants were undergraduate students which limits the generalizability of the findings. Although it seems reasonable to assume similar patterns would be detected among other populations, future research would be required to make this firm conclusion.

Conclusion

There are different ways of assessing people's well-being. Thinking about and reflecting on how your life is going at one time involves a cognitive process that differs considerably from how your life is going in real time (or near real time). Global reports were consistently higher than aggregated daily states. This has several important implications for the interpretation of results that rely on these different methods. In sum, life is considered more satisfying, more meaningful, and is characterized to a greater extent by more intense positive and negative emotions when reflecting on life in general than when reflecting on daily life in real time.

Acknowledgements

References

Affleck G, Zautra A, Tennen H, & Armeli S (1999). Multilevel daily process designs for consulting and clinical psychology: A preface for the perplexed. Journal of Consulting and Clinical Psychology, 67, 746–754. 10.1037/0022-006X.67.5.746 [PubMed: 10535241]

Amrhein V, Greenland S, & McShane B (2019). Scientists rise up against statistical significance. Nature, 567, 305–307. [PubMed: 30894741]

Bauer DJ, Preacher KJ, & Gil KM (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. Psychological Methods, 11(2), 142–163. [PubMed: 16784335]

Bless H, & Schwarz N (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. Advances in Experimental Social Psychology, 42, 319–374.

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

Bradburn NM, Rips LJ, & Shevell SK (1987). Answering autobiographical questions: The impact of memory and inference on surveys. Science, 236(4798), 157–161. [PubMed: 3563494]

Brandstätter H (2007). The time sampling diary (TSD) of emotional experience in everyday life situations. In Allen J & Coan J (Eds.), The handbook of emotion elicitation and assessment (pp. 318–331). New York, NY: Oxford University Press.

Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, & Stone AA (2008). The accuracy of pain and fatigue items across different reporting periods. Pain, 139(1), 146–157. [PubMed: 18455312]

Conner TS, Feldman Barrett L (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. Psychosomatic Medicine, 74, 327–337. [PubMed: 22582330]

Deaton A, & Stone AA (2016). Understanding context effects for a measure of life evaluation: How responses matter. Oxford Economic Papers. 10.1093/oep/gpw022

Diener E, Emmons RA, Larsen RJ, & Griffin S (1985). The satisfaction with life scale. Journal of Personality Assessment, 49(1), 71–75. [PubMed: 16367493]

Diener E, Lucas RE, & Oishi S (2018). Advances and open questions in the science of subjective well-being. Collabra: Psychology, 4(1), 1–49. [PubMed: 30637365]

Diener E, Suh EM, Lucas RE, & Smith HL (1999). Subjective well-being: Three decades of progress. Psychological Bulletin, 125(2), 276–302.

Faul F, Erdfelder E, Buchner A, & Lang AG (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. [PubMed: 19897823]

Feldman Barrett L, & Russell JA (1998). Independence and bipolarity in the structure of current affect. Journal of Personality and Social Psychology, 74, 967–984. 10.1037/0022-3514.74.4.967.

Fleeson W (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. Journal of Personality and Social Psychology, 80(6), 1011–1027. [PubMed: 11414368]

Fleeson W, & Jayawickreme E (2015). Whole trait theory. Journal of Research in Personality, 56, 82–92. [PubMed: 26097268]

Goetz T, Bieg M, Lüdtke O, Pekrun R, & Hall NC (2013). Do girls really experience more anxiety in mathematics? Psychological Science, 24(10), 2079–2087. [PubMed: 23985576]

Heintzelman SJ, & King LA (2014). Life is pretty meaningful. The American Psychologist, 69(6), 561–574. [PubMed: 24490907]

Henderson LW, Knight T, & Richardson B (2013). An exploration of the well-being benefits of hedonic and eudaimonic behaviour. The Journal of Positive Psychology, 8(4), 322–336.

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: Russell Sage.

Kahneman D & Riis J (2005). Living, and thinking about it: Two perspectives on life. In Huppert FA, Baylis N, & Keverne B (Eds.), The science of well-being. (pp. 285–304). Oxford: Oxford University Press.

Kopperud KH, & Vitterso J (2008). Distinctions between hedonic and eudaimonic well-being: Results from a day reconstruction study among Norwegian jobholders. The Journal of Positive Psychology, 3(3), 174–181.

Maas CJM, & Hox JJ (2005). Sufficient sample sizes for multilevel modeling. Methodology, 1, 86–92. 10.1027/1614-1881.1.3.86.

McFarland C, Ross M, & DeCourville N (1989). Women's theories of menstruation and biases in recall of menstrual symptoms. Journal of Personality and Social Psychology, 57(3), 522–531. [PubMed: 2778636]

Meade AW, & Craig SB (2012). Identifying careless responses in survey data. Psychological Methods, 17(3), 437–455. [PubMed: 22506584]

Newman DB, Nezlek JB, & Thrash TM (2018). The dynamics of searching for meaning and presence of meaning in daily life. Journal of Personality, 86(3), 368–379. 10.1111/jopy.12321 [PubMed: 28423186]

Newman DB, Sachs ME, Stone AA, & Schwarz N (in press). Nostalgia and well-being in daily life: An ecological validity perspective. Journal of Personality and Social Psychology. 10.1037/pspp0000236

Newman DB, Schwarz N, Graham J, & Stone AA (2019). Conservatives report greater meaning in life than liberals. Social Psychological and Personality Science, 10(4), 494–503. 10.1177/1948550618768241 [PubMed: 31565182]

Nezlek JB (2005). Distinguishing affective and non-affective reactions to daily events. Journal of Personality, 73(6), 1539–1568. [PubMed: 16274445]

Nezlek JB (2012). Diary methods for personality and social psychology (Sage library of methods in social and personality psychology). London, England: Sage.

Nezlek JB (2017). A practical guide to understanding reliability in studies of within-person variability. Journal of Research in Personality, 69, 149–155.

Nezlek JB, Newman DB, & Thrash TM (2017). A daily diary study of relationships between feelings of gratitude and well-being. The Journal of Positive Psychology, 12(4), 323–332.

O'Keefe DJ (2007). Brief report: Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. Communication Methods and Measures, 1(4), 291–299.

Raudenbush SW, Bryk AS, & Congdon R (2011). HLM 7 for Windows [Computer software]. Skokie, IL: Scientific Software International.

Ross M (1989). Relation of implicit theories to the construction of personal histories. Psychological Review. 10.1037//0033-295x.96.2.341

Schimmack U, Diener E, & Oishi S (2002). Life-satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. Journal of Personality, 70(3), 345–384. [PubMed: 12049164]

Schwarz N (2012). Why researchers should think "real-time": A cognitive rationale. In Mehl MR & Conner TS (Eds.), Handbook of research methods for studying daily life (pp. 22–42). New York: Guilford.

Schwarz N, & Clore GL (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. Journal of Personality and Social Psychology. 10.1037//0022-3514.45.3.513

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: Russell-Sage.

Schwarz N, Strack F, & Mai H-P (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. Public Opinion Quarterly. 10.1086/269239

Schwarz N, & Xu J (2011). Why don't we learn from poor choices? The consistency of expectation, choice, and memory clouds the lessons of experience. Journal of Consumer Psychology, 21(2), 142–145.

Shiffman S, Stone AA, & Hufford MR (2008). Ecological momentary assessment. Annual Review of Clinical Psychology, 4, 1–32.

Steger MF, Frazier P, Oishi S, & Kaler M (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. Journal of Counseling Psychology, 53(1), 80–93.

Steptoe A, Deaton A, & Stone AA (2015). Subjective wellbeing, health, and ageing. The Lancet, 385, 640–648. 10.1016/S0140-6736(13)61489-0

Stone AA, Broderick JE, & Schwartz JE (2010). Validity of average, minimum, and maximum end-of-day recall assessments of pain and fatigue. Contemporary Clinical Trials, 31(5), 483–490. [PubMed: 20620239]

Stone AA, Broderick JE, Shiffman SS, & Schwartz JE (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(1–2), 61–69. [PubMed: 14715390]

Stone AA, Schwartz JE, Neale JM, Shiffman S, Marco CA, Hickcox M, … Cruise LJ (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. Journal of Personality and Social Psychology, 74(6), 1670–1680. [PubMed: 9654765]

Tennen H, Suls J, & Affleck G (1991). Personality and daily experience: The promise and the challenge. Journal of Personality, 59, 313–338.

Tov W, & Lee HW (2016). A closer look at the hedonics of everyday meaning and satisfaction. Journal of Personality and Social Psychology, 111(4), 585–609. [PubMed: 26619303]

Waytz A, Hershfield HE, & Tamir DI (2015). Mental simulation and meaning in life. Journal of Personality and Social Psychology, 108(2), 336–355. [PubMed: 25603379]

Wirtz D, Kruger J, Napa Scollon C, & Diener E (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. Psychological Science, 14(5), 520–524. [PubMed: 12930487]

Table 1.

Descriptive statistics of daily and global reports.

| Variable | | **Daily Reports** | | | | **Global Reports** | | |
| | Intercept | **Variation** | | Reliability | Validity | M | SD | Reliability |
		Within	Between					
Positive activated affect	3.76	1.27	1.32	.84	.72	4.46	1.29	.91
Positive deactivated affect	3.54	1.06	.98	.82	.71	3.96	1.14	.86
Negative activated affect	3.06	1.18	.72	.61	.60	3.66	1.16	.79
Negative deactivated affect	2.27	1.07	.82	.79	.68	2.86	1.26	.87
Satisfaction with life	4.57	1.54	.92	.79	.71	4.87	1.41	.78
Meaning in life – presence	4.20	1.20	1.17	.78	.67	4.75	1.36	.82
Meaning in life – search	3.45	1.22	1.22	.84	.33	4.98	1.35	.91
Nostalgia	2.76	1.54	1.54	.89	.62	3.88	1.58	.91

Table 2.

Comparisons of global reports and aggregated daily state reports.

Variable	Daily report averages		Global reports		Main effect				Interaction			
	Global report first group	Diary first group	Global report first group	Diary first group	b	t	p	d	b	t	p	d
Positive activated affect	3.71	3.81	4.47	4.46	.71	11.31	<.001	.72	.11	.90	.369	.11
Positive deactivated affect	3.45	3.64	3.99	3.93	.41	7.46	<.001	.48	.25	2.29	.023	.29
Negative activated affect	3.00	3.13	3.81	3.51	.60	9.44	<.001	.60	.43	3.46	.001	.44
Negative deactivated affect	2.17	2.38	3.01	2.71	.58	9.46	<.001	.60	.50	4.14	<.001	.54
Satisfaction with life	4.61	4.53	4.76	4.99	.30	4.36	<.001	.28	−.32	2.38	.018	.30
Meaning in life – presence	4.17	4.23	4.78	4.72	.55	8.03	<.001	.52	.10	.76	.446	.09
Meaning in life – search	3.29	3.62	5.15	4.78	1.51	15.59	<.001	1.00	.71	3.77	<.001	.48
Nostalgia	2.67	2.85	4.31	3.44	1.12	13.28	<.001	.85	1.04	6.77	<.001	.87

Note: The main effect refers to the difference between global reports and aggregated daily states collapsed across condition. Because reliable effect sizes could not be calculated for the specific multilevel models that used difference scores, we calculated Cohen's d effect sizes from repeated measures ANOVA models.

Table 3.

Correlations between global reports and 1) peak states, 2) last three daily reports, and 3) daily averages across conditions.

Variable	Correlation between peak (maximum) and global report					Correlation between last three days and global report					Correlation between daily average and global report				
	Global first	Diary first	z	p	q	Global first	Diary first	z	p	Q	Global first	Diary first	z	p	q
Positive activated affect	.53 [.39, .65]	.69 [.58, .77]	1.98	.048	.257	.63 [.51, .73]	.61 [.48, .71]	.25	.802	.033	.68 [.58, .77]	.70 [.60, .78]	.25	.806	.032
Positive deactivated affect	.45 [.30, .58]	.58 [.44, .68]	1.31	.189	.170	.55 [.42, .67]	.62 [.50, .72]	.78	.434	.101	.66 [.55, .75]	.73 [.63, .80]	1.05	.293	.136
Negative activated affect	.56 [.43, .67]	.40 [.24, .54]	1.61	.107	.209	.45 [.29, .58]	.49 [.34, .62]	.44	.662	.057	.60 [.47, .70]	.56 [.43, .67]	.43	.671	.055
Negative deactivated affect	.63 [.51, .73]	.48 [.33, .61]	1.72	.086	.222	.52 [.37, .64]	.57 [.44, .68]	.59	.559	.076	.63 [.51, .73]	.70 [.60, .78]	1.00	.315	.130
Satisfaction with life	.34 [.17, .49]	.63 [.51, .72]	2.98	.003	.386	.47 [.32, .60]	.63 [.50, .72]	1.75	.079	.227	.57 [.44, .68]	.76 [.67, .82]	2.64	.008	.342
Meaning in life – presence	.44 [.29, .58]	.62 [.50, .72]	1.98	.048	.256	.52 [.38, .64]	.59 [.46, .69]	.68	.494	.089	.64 [.52, .73]	.66 [.55, .75]	.30	.765	.039
Meaning in life – search	.22 [.05, .39]	.46 [.31, .59]	2.10	.036	.272	.09 [−.09, .27]	.45 [.30, .58]	3.06	.002	.396	.20 [.02, .36]	.49 [.34, .61]	2.57	.010	.333
Nostalgia	.49 [.35, .62]	.61 [.49, .71]	1.32	.187	.171	.44 [.28, .57]	.61 [.49, .71]	1.86	.063	.241	.52 [.38, .64]	.76 [.67, .83]	3.28	.001	.425

Note: Comparisons of correlations were conducted with a Fisher's r to z transformation and Cohen's q is provided as an effect size estimate. 95% confidence intervals of the correlations are listed in brackets.