Measuring Experiential Well-Being among Older Adults

Richard E. Lucas[1], **Vicki A. Freedman**[2], and **Deborah Carr**[3]

[1]Department of Psychology, Michigan State University

[2]Institute for Social Research, University of Michigan

[3]Department of Sociology, Boston University

Abstract

Experienced well-being measures tap a distinct form of subjective well-being (SWB) and have different age-related properties than the more widely studied evaluations of life satisfaction. Unlike evaluations of the quality of life as a whole, experiential measures capture affective reactions soon after they occur. Recent advances in measurement have allowed for the inclusion of such experiential measures even in large-scale studies. However, respondent burden remains a concern; hence, surveys have also employed shorter experiential modules. The psychometric properties of these brief measures are not well understood. We examine the psychometric characteristics, including the factor structure and correlations with theoretically relevant criteria, of experienced wellbeing measures included in two supplements to the Panel Study of Income Dynamics (PSID). The first supplement included a detailed time diary whereas the second included a brief review of the prior day. Results show that for the detailed time diaries a single index of affective experience provides a useful summary of the associations among individual affect items, both within and between participants. For the abbreviated method, two or more subscales better describe the underlying structure.

Keywords

Subjective well-being; measurement; affect

Subjective well-being (SWB) reflects a person's own assessment of the quality of his or her life and experiences (Diener, Lucas, Schimmack, & Helliwell, 2009; Diener, Suh, Lucas, & Smith, 1999). These subjective evaluations are important because, unlike objective measures of material goods or physical health, they reflect individuals' own perceptions of their lives. SWB measures are valuable both for understanding individual-level processes and for broader policy purposes, as they can be used to quickly and efficiently assess the impact of decisions, life circumstances, and policies that may have a broad range of primary and incidental effects on quality of life (Diener et al., 2009).

Historically, researchers have assessed SWB using broad, global measures. For instance, many assessment tools ask respondents to rate their satisfaction with life or their typical

Correspondence: concerning this article should be addressed to Richard E. Lucas, 316 Physics Rd., Michigan State University, East Lansing, MI 48823. lucasri@msu.edu.

levels of positive emotions using simple, face-valid items (e.g., Diener, Emmons, Larsen, & Griffin, 1985; Lucas, Diener, & Larsen, 2003). These measures have a number of advantages: they are simple and efficient, they are generally well understood by most respondents, and they typically have high levels of reliability and validity (Cheung & Lucas, 2014; Diener et al., 2009). They also allow individuals to consider diverse aspects of their lives (e.g., health, finances, social activity, mental state), while also allowing for idiosyncratic weighting of the importance of those domains across individuals (e.g., people may vary in the value they place on high incomes). Much of what researchers know about SWB come from global, self-report measures (Diener et al., 1999).

Nevertheless, global measures have limitations. One concern is that the process of evaluating one's life is a complex task that requires respondents to recall, evaluate, and weight the importance of a variety of domains, experiences, and events from their lives. This task may be cognitively taxing for respondents, who may not be willing or able to comprehensively evaluate all the life domains that should be relevant. Research on reports of affect over different time frames shows that when respondents are asked to think about their affective experiences over a very long time, they respond more quickly than would be expected if they were actually searching their memory for relevant information (Robinson & Clore, 2002). Instead, respondents may rely on a variety of heuristics or mental shortcuts that can affect the validity of the measures (Kahneman, 1999; Schwarz & Strack, 1999).

Partly in response to these types of concerns, researchers have suggested alternative approaches to assessing SWB that do not rely on global judgments. For instance, Kahneman (1999) argued that simple judgments about whether a person is feeling positive or negative *at a single moment* can be made quite easily and that such judgments are unlikely to be affected by comparison standards or other undesirable context effects. Thus, he argued, it would be possible to obtain a measure of *objective happiness* by repeatedly sampling people's affective reactions using experience sampling or ecological momentary assessment techniques (Stone, Shiffman, & DeVries, 1999).

The main limitation of this alternative approach is that it is burdensome, both for researchers and respondents. To obtain momentary data, respondents must repeatedly be interrupted throughout their day—for instance, by mobile telephones or other handheld devices—to respond to questions about their feelings and activities, questions that may feel intrusive and repetitive after multiple administrations. In response to this concern about respondent burden, researchers developed alternative approaches that address the potential limitations of global measures, while solving the problems of respondent burden that are associated with momentary assessment techniques. These techniques typically involve short-term retrospective evaluations of a single day, using methodologies borrowed from research on time use (Kahneman & Krueger, 2006; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004; Krueger et al., 2009; Lee, Hofferth, Flood, & Fisher, 2016).

Although time-use-based approaches for assessing affective reactions to situations and events have been used for decades (see Lee et al., 2016, for a review), the introduction of the Day Reconstruction Method (DRM) by Kahneman et al. (2004) led to an increase in the use of these techniques for assessing SWB. In the DRM, participants are asked to reconstruct the

prior 24-hour period in terms of distinct "episodes" in which the respondent participated from waking to the time he or she went to sleep at night. For instance, the respondent may indicate that she woke at 6 AM and ate breakfast alone for the next forty-five minutes. After listing all episodes for the entire day, respondents are then asked to answer a series of questions about their affective experiences during each episode (the original DRM design asked respondents about all episodes; later implementations sample from the full set and ask detailed questions about this subsample). These affective ratings can be aggregated across the day to get an overall measure of affective experience, or the individual episodes can be analyzed to test questions about the within-person effects of situations.

The DRM technique can be used effectively by phone to allow participants to reconstruct the prior day, along with the emotions that were experienced throughout that day. For instance, among older couples in the Panel Study of Income Dynamics, 76% of joint activities mentioned by one spouse were found within an hour window on the other spouse's time diary (Freedman, Stafford, Conrad, Schwarz, & Cornman, 2012). However, the full time diary still takes approximately 20–25 minutes to complete. Consequently, investigators have developed additional briefer methods of recalling key activities on the prior day and associated emotions. For instance, in a supplement to the Health and Retirement Study, a narrower set of activities was asked about (Smith, Ryan, Queen, Becker, & Gonzalez, 2014). More recently, the Panel Study of Income Dynamics developed a mixed mode (web/paper) supplement on wellbeing that includes an abbreviated DRM (which, in this paper, we refer to as the "Abridged DRM") that collects the highlights of the prior day and then asks respondents to assess how much of the day they felt various emotions. This abbreviated version takes approximately five minutes to complete in a web application.

Initial research on the DRM and related methods focused primarily on the extent to which various situational characteristics of the episodes that respondents described correlated with the affect that they reported (e.g., Kahneman et al., 2004). As research has progressed, however, studies have focused on more nuanced psychometric questions. For instance, recent work has examined the year-to-year stability of DRM-based measures of SWB (Hudson, Lucas, & Donnellan, 2017), the validity of DRM-based measures in a large German sample (Anusic, Lucas, & Donnellan, 2017), and the factor structure of affect items used in the American Time Use Survey (ATUS)(Lee et al., 2016).

The latter paper by Lee et al. (2016) is especially relevant to our study because it addresses questions about how various affect items that are assessed within the context of a DRM survey should be combined. Global measures of affect and emotion are supported by a long tradition of factor-analytic evidence for the structure of these ratings (e.g., Watson, Clark, & Tellegen, 1988). This evidence guides how different emotion items are combined to create overall indexes of emotional experience. For instance, Watson et al. (1988) showed that positive and negative affect items can be meaningfully summarized using separable positive and negative affect dimensions (see Schimmack, 2008, for a review). However, DRM measures present unique psychometric challenges for researchers who need to decide whether and how to combine items into composite scores. First, DRM measures often include relatively small numbers of items to minimize respondent burden. Prior work on the

structure of longer lists of affect items cannot always be used to inform decisions with the relatively small set of items included in DRM studies.

Perhaps more importantly, the data that result from DRM methodology (data that include multiple activities for each person in the study) have a multilevel structure, and the associations among items may differ at each of these levels. Thus, as Lee et al. (2016) noted, one can factor analyze between-person or within-person associations among affect items, and different measurement decisions may result at each of the two levels. Lee et al. (2016) used such a technique and identified two factors both in within- and between-person data from the ATUS. However, these results may be unique to the specific set of items included in the ATUS, and therefore, it will be useful to determine whether similar structures emerge with different item sets in different samples.

The goal of the current study is to examine the psychometric properties of two experienced wellbeing measures collected from older adults in the Panel Study of Income Dynamics: the full DRM and an abridged DRM that asks just once about respondents' experiences over the entire course of the day. For the former, we use factor analytic techniques that are appropriate for the multilevel structure of these data. In addition, we examine the stability of the derived variables from weekday to weekend and assess associations with various theoretically relevant criterion variables. These analyses will not only contribute new knowledge about the psychometric properties of DRM-based affect measures, they will provide critical practical guidance for researchers who wish to use the measures from the publicly available and widely analyzed PSID study.

Sample and Methods

We focus primarily on the 2013 wave of the DUST supplement to the PSID. The PSID is the world's longest running national observational household panel study. Begun in 1968 with approximately 5,000 families, interviews have been conducted annually through 1997 and every two years by telephone since then with one household informant. When adult children leave home and form their own economically independent households, they become eligible for the study; consequently the sample grows naturally. Approximately 9,100 interviews were conducted in 2013. In addition, in 2013, a supplemental study of adults ages 60 and older in the PSID and their spouses or partners was conducted by telephone to obtain detailed information on disability, use of time, and experienced wellbeing.[1] 1,776 respondents participated in all (72% of eligible households completed at least one interview).

In 2016, the PSID administered an additional supplement, the Wellbeing and Daily Life supplement, which includes a variety of additional questions on wellbeing, including the abbreviated experienced wellbeing measure. The 1,365 participants who participated in the DUST and also completed the Wellbeing and Daily Life Supplement were included in additional analyses involving this brief measure. Detailed information about all measures

---

[1] A similar time-use component was also administered in 2009, though with a smaller, more select sample. Because the 2013 wave included more participants, that 2013 wave is the focus of the current study.

used in the PSID and DUST supplement are available at: https://psidonline.isr.umich.edu/guide/documents.aspx

The main PSID interview consists of questions about a variety of topics including income, program participation, employment, housing, expenditures, health, wealth, family composition, and education. The DUST supplement also includes a DRM module. Participants were administered a time diary covering the prior 24 hour period (including what they were doing, how long it took, where they were, and who they were with). Three activities were then selected for more detailed assessment of emotions and feelings ("calm", "happy", "frustrated", "worried", "sad"[2]) experienced during those activities. Participants completed this time diary up to two times, once for a weekday and once for a weekend. In order to maximize waking hours, one activity was randomly selected from 8 AM to 11:59 AM, a second from 12:00 PM to 3:59 PM, and a third from 4:00 PM to 8:00 PM. Thus, we have up to six activity-level reports per person.

The 2016 wellbeing module used an abbreviated daily emotion report that first asked participants to report what time they woke up, what time they went to sleep, whether they spent any time in the types of activities that are assessed during the full DRM, and who they spent time with on the prior day. Then, participants were asked to report on their overall levels of various emotions using the following question stem: "Thinking about the whole day yesterday, how much of the day did you feel each of the following?" Respondents were then asked to respond to a list of 12 adjectives: The five from the DUST, plus "enthusiastic," "content," "interested," "angry," "stressed," "lonely", and "bored." Participants responded on 5-point scale with response options ranging from "None of the Day" to "All of the Day."

In addition to examining the associations among the various well-being items, we also examined validity by testing the associations between constructed affect scale scores and additional theoretically relevant criteria. These criteria include measures of a variety of constructs that have been shown to be associated with well-being in the past. For instance, participants were asked a series of eight dichotomous questions about health problems that limited their activities during the past seven days. Specifically, they were asked whether any of these problems limited their activities, and if they responded "Yes", they were asked how many days their activities were limited. These eight dichotomous items were summed to get an overall score.

Participants were also asked six dichotomous questions about difficulties that they had, including difficulties with hearing; seeing, even with glasses; concentrating, remembering, or making decisions because of a physical, mental, or emotional condition; walking or climbing stairs; dressing or bathing; and doing errands alone (due to a physical, mental, or emotional condition). Again, these dichotomous items were summed to create an overall measure of difficulties. Three additional items were used to assess participants' opinions about their memory.

---

[2]·The items "pain" and "tired" were also included in the survey, but are not analyzed here because they are considered somatic symptoms rather than emotions.

A 15-item measure of the Big Five personality traits that had been developed for use in the German Socio-Economic Panel Study was also assessed. In this sample, alphas for Extraversion, Neuroticism, Conscientiousness, Agreeableness, and Openness were: 0.52, 0.64, 0.54, 0.47, 0.65. Although somewhat low, these values are consistent with those reported for short-form measures in other large-scale panel studies (Donnellan & Lucas, 2008). Participants also completed a measure of self-efficacy consisting of five dichotomous items. The alpha for this measure was 0.55. In addition, participants completed a four-item spirituality scale with a reliability of 0.87.

Finally, participants were also asked a series of questions about the quality of their relationships with their family and (if they had one) their spouse or romantic partner (e.g., how much these people appreciated the respondent, how much they argue). The reliabilities for these sets of items were 0.81 for romantic relationships and 0.49 for family relationships. Participants also indicated in a single item how many friends they had.

Analytic Approach

The primary focus of this study is on the associations among the emotion items, taking into account the multilevel structure of the data when appropriate. Specifically, we use multilevel factor analytic techniques implemented in Mplus 7.1 to isolate within-person associations from between-person associations. We first describe these associations using correlation coefficients and then factor analyze these associations separately for each level. We make no predictions about differences in associations across levels, yet this approach allows for different structures to emerge. When evaluating the factor analytic results, we assess whether there is clear evidence for factor structures that mirror what has been found in global reports of affect. We then compare the results from the multilevel DRM data to a factor analysis of the longer list of emotion/feeling items from the brief wellbeing module. Finally, we assess the psychometric properties of the scale scores that are suggested from these analyses by examining stability over time and associations with other constructs.

Results

Means and standard deviations for the emotion items used across the two study components are presented in Table 1. This table is separated by level of aggregation, with the left columns reflecting activity-level data, the middle columns reflecting person-level data, and the farthest right columns reflecting the longer set of items included in the abridged DRM (note the difference in response options across the two methods). The means for the two levels of analysis from the full DRM will be quite similar, as differences are due to slight variations in the number of activity reports that contribute to each person's aggregated score. Standard deviations will differ across levels, however, depending on the amount of within-person variance in these reports.

Correlations among full-DRM item scores, separated by level of analysis, are presented in Table 2. The correlations above the diagonal reflect within-person associations among the five adjective ratings; the correlations below the diagonal reflect the between-person associations among each adjective aggregated across all ratings provided by an individual for up to six activities. As would be expected, the correlations are stronger for the

aggregated correlations, as these remove considerable amounts of measurement error. At the between-person level, all correlations are moderate to large. Correlations between items of the same valence tend to be slightly higher than correlations between items with different valences, though this is not always the case. At the within-person level, this same pattern emerges, though with weaker correlations overall. Consistent with the results from the full DRM data, correlations between items from the abridged DRM (not shown) were moderate to strong, and correlations among same-valenced items tended to be larger than correlations between two items of opposite valence (though again, this pattern does not always hold).

Next, we conducted factor analyses separately for the full and abridged DRM data. The scree plots in Figure 1 show the results from simultaneous exploratory factor analyses conducted on all five items from the full DRM at the between- (left panel) and within-person (right panel) levels of analysis. As would be expected due to differences in the amount of measurement error at each level of analysis, more variance is explained by the first factor in the between-person analyses than for the within-person analyses. The scree plot makes a strong case for a single factor model, as there is only one eigenvalue greater than one in both the between- and within-person analyses. Although there is a slight break in the scree plot after two factors, the break after the first factor is much clearer. This is true both in the between-person and within-person analyses. Thus, although the pattern of correlations and factor analytic results suggest that positive and negative emotion items might be separable, there is stronger evidence that combining all five emotion items into a single scale score is reasonable.

We next conducted similar analyses with the 12 items from the abridged DRM in the Wellbeing Module. As the scree plot in the left panel of Figure 2 shows, there was again a very strong first factor, along with two additional factors with eigenvalues greater than 1. After this third factor, there was a relatively clear break, suggesting that at most, three factors should be retained. Factor loadings for the three-factor solution are presented in Table 3. One factor is defined by the positive items, a second factor is defined by all negative items excluding lonely and bored, and a third factor is defined by these last two negative items.

Finally, for the most direct comparison of the full DRM methodology with the abridged method used in the wellbeing module, we conducted a third factor analysis that focused only on the five overlapping emotion items. The second panel of Figure 2 shows that there was again a strong first factor. Only one factor had an eigenvalue above 1, and there was a clear break in the screen plot after this factor. Thus, consistent with the analysis of the five items from the full DRM component, it appears that these items are best summarized with a single wellbeing score.

Associations Among Different Wellbeing Measures

Table 4 reports the correlations between each of the three scale scores that can be constructed from the five items from the full DRM, separately for the weekday and weekend reports. These correlations are useful for three purposes. First, within each day, they clarify how strongly the narrower constructs correlate with one another and with the aggregate emotion measure constructed from all emotion items. This analysis provides further

evidence about the extent to which it is justifiable to combine positive and negative items or whether they should be kept separate. For instance, the positive and negative affect scale scores correlate relatively strongly with one another ($r = -0.57$ and -0.55), and extremely strongly with the composite measure. Especially when considered in light of the somewhat low reliabilities of the valenced sub-scales, this pattern of correlations argues for their combination.

Second, the cross-time correlations provide evidence about the extent to which one day's worth of experiential measures provide a reasonable indicator of one's overall emotional well-being. One could speculate that emotions might fluctuate from day to day, which would mean that it would take multiple days of assessment to get a reasonably stable estimate of average well-being (though see Hudson et al., 2017). The results presented in Table 4 suggest, however, that about 50% of the variance in these measures is variance that is stable across a single weekday and a single weekend day.

Third, the measure-specific cross-time correlations (e.g., the correlation between weekday positive affect and weekend positive affect) tend to be slightly stronger than the cross-measure cross-time correlations. Using the logic of the multitrait-multimethod matrix (Campbell & Fiske, 1959), this suggests that there is a small amount of construct-specific variance in the narrow scale-score measures. Thus, although these distinct positive and negative affect subscales capture a considerable amount of variance that is shared across items, the narrow measures could potentially correlate differentially with additional predictors and outcomes.

Table 5 shows the correlations between each of the three scale scores derived from the full DRM data (wellbeing, positive affect, and negative affect; all items are aggregated across weekend and weekday assessments for this analysis) and seven scale scores that could be constructed using the data from the abridged DRM. In the first three rows, correlations between full and abridged DRM-based scale scores using the same five items are presented. Rows four through seven show correlations with abridged-DRM scale scores constructed from all items, using the three-factor solution as a guide. Thus, the fourth row includes all 12 items; the fifth row includes the 5 positive items, the sixth row includes all negative items excluding lonely and bored; and the seventh row includes scale scores constructed just from these final two items.

Again, a few clear patterns emerge. First, when focusing on the associations between scale scores constructed from the same items, correlations are moderate in size (rs = 0.41, 0.38, 0.36). Second, these correlations do not change when the additional items from the abridged DRM scale are added; the correlations between full-and abridged measures of overall wellbeing, positive affect, and negative affect (excluding boredom and loneliness) are identical, regardless of whether the full set of items from the abridged method is included. Finally, there is some, but not strong, evidence for discriminant validity of the distinct positive and negative affect measures, as the correlations between two positive affect measures or two negative affect measures assessed using different methods are only slightly higher than cross-construct, cross-method correlations. Again, however, the difference in these correlations is not large, suggesting only minimal discriminant validity.

Correlations with Predictors and Outcomes

As a final examination of the impact of different measurement approaches and decisions decisions on the validity of the affect measures, we examined the association between each of the scale scores (the combined well-being measure, along with the positive and negative affect measures) and a variety of outcomes that are theoretically related to well-being and that were also assessed in the DUST or PSID. The idea behind these analyses is that if the narrower subscales contained important unique variance, they would show distinct patterns of correlations with criterion variables.

Table 6 shows the correlations between the three DRM-based measures and five different assessments of a standard single-item life satisfaction scale, three from the core PSID and two from the 2013 DUST (one at the beginning of the survey and one at the end). The correlations between positive affect and the various life satisfaction measures are similar to and sometimes slightly higher than the correlation between negative affect and these measures. In addition, the correlation with between the life satisfaction measures and the aggregated emotional well-being measure is generally the same size or larger as the correlation with positive affect. Correlations from the full-DRM measures are very similar in size to those obtained from the abridged DRM.

Table 7 shows the correlations between the full and abridged DRM measures and the Big Five personality traits, along with a measure of self-efficacy and spirituality. Correlations are generally quite similar across the various measures, even for the theoretically relevant associations between extraversion and positive affect and between neuroticism and negative affect. Thus, there is not a great deal of discrimination across the various measures when evaluating correlations with personality and other individual difference measures.

Table 8 shows the correlations with a set of variables that tap features of one's social relationships (including evaluations of the quality of respondents' relationships with their families and romantic partners, along with a more objective measure of the number of friends they have). Again, the correlations (which are small to moderate in size) are relatively similar across measures.

Finally, Table 9 reports associations with a number of variables that index health and disability status. Again, correlations with the affect measures are small to moderate in size, and quite similar across all measures regardless of whether they were measured using the full or abridged DRM.

Discussion

Subjective well-being measures are designed to assess people's overall evaluation of the quality of their lives as a whole, yet there are multiple ways such evaluations can be obtained. Respondents can be asked to reflect on their lives and provide a judgment of their overall satisfaction with the various features and domains. Alternatively, researchers can track and aggregate the emotional responses that people have to their lives. Reports of short-term affective reactions provide an intuitively appealing alternative to global judgments of well-being, yet questions remain about their psychometric properties. Specifically, because

respondent burden is a concern when experiential measures are used (especially in the context of large-scale surveys where respondents are completing many different measures) relatively brief, ad hoc lists of emotions are often used in studies that employ the DRM. The extent to which these items can and should be combined into broader aggregates is in question, and the psychometric properties of these measures when used to track feelings within and between people over time are unclear. The goal of the current project was to assess these issues in a sample of older adults assessed as part of a large-scale panel study. A secondary goal was to examine whether a simpler abridged-DRM approach to understanding experienced affect throughout a day provides similar information to a more intensive experiential approach that asked respondents about specific episodes of their day.

In regard to the structure of the items assessed, our results showed that both within and between persons, associations among the specific emotion items included in the full DRM measure were moderate to strong. Importantly, even emotion items of different valence tended to correlate at least moderately with one another, and both the pattern of correlations and factor analyses of these correlations suggested that these items might best be combined into a single scale score. This conclusion was further supported by the pattern of associations with additional criteria: Aggregated positive and negative affect scales tended to show similar correlations with these criteria, and in general, a broader aggregate measure that averaged across all positive and negative items showed the strongest associations with criteria. Thus, for this specific set of items in this population, there was not strong justification for examining separate positive and negative affect scales.

Additional analyses using an abridged DRM measure that only asked respondents to describe the amount of time they spent in various activities and then to provide an overall assessment of their affective experience over the course of the day showed similar results. When the same items from the DRM were analyzed, the same single-factor structure received support, separate positive and negative affect measures were moderately correlated, and patterns of correlations were similar for the various measures that could be constructed. These results suggest that the specific factor structure that emerged in the analyses of the full DRM data did not result from the unique methodological features of the this approach, but instead reflects the typical associations that would be found among these specific affect items. In other words, researchers interested in the DRM method solely for measuring aggregated daily wellbeing (and not time use or moment-to-moment changes) may find the abridged method to be adequate. If, however, both time use and wellbeing are of interest, the full DRM may be preferred.

The use of the abridged DRM does allow for a longer list of affect items to be included, and when we analyzed the full set of 12 items, a slightly different factor structure emerged. Specifically, although there was a clear and strong first factor, there was more evidence that a three-factor structure was needed to adequately account for the associations among items. Importantly, the subscales that could be constructed based on this three-factor solution were also quite strongly correlated with one another, and there was little evidence of discriminant validity when examining the correlations across methods or when examining correlations with external criteria. However, if there are strong theoretical reasons to focus on separable positive and negative affect measures, the longer set of items included in the abridged DRM

may allow for greater discrimination between these constructs. For example, studies based on the DUST exploring marital dynamics reveal distinctive effects where spousal caregiving is linked to wives' reports of happiness (Freedman, Cornman, & Carr, 2014), whereas the receipt of marital support is associated with husbands' reports of heightened frustration (Carr, Cornman, & Freedman, 2017). Thus, researchers' decisions regarding which measures to use can be informed by specific theoretical links between emotion constructs and predictors of interest.

One question that emerges from these results is why clearly separable positive and negative affect dimensions did not emerge in these data, given that a two dimensional structure is often found in factor analyses of emotion ratings (Schimmack, 2008). The size of the association between any two emotion items depends on factors beyond the valence of those items. For instance, independent positive and negative affect dimensions are often conceptualized as rotational variants of two different underlying dimensions, arousal and valence. Two items that have strong loadings on arousal may exhibit weak or even positive correlations, even if they differ on valence. In contrast, two items that have moderate loadings on arousal, and opposite loadings on the valence dimension would be expected to show strong negative associations. Thus, the expected correlation between aggregated positive and negative affect measures will depend on the specific items that are included. Many factors can guide item selection, and the specific items included in the DUST represent a reasonable set of prototypical emotion terms. However, this particular set may not cleanly reflect the dimensional structure that has been found in past factor analytic work. This is why psychometric work like that included in this study is needed. In addition, the time frame over which respondents are asked to reflect can impact associations between oppositely valenced affect items (Schimmack, 2008), and the weaker evidence for distinct positive and negative affect dimensions may result from the focus on in-the-moment experienced affect.

It is important to acknowledge, however, that this study does have some limitations that impact the conclusions that can be drawn. First, the population sampled included older adults. Research shows that older adults' emotional experience may differ in important ways from younger adults' experiences (Carstensen, Fung, & Charles, 2003). Thus, conclusions about the structure of these specific items should be constrained to this population. Second, although the full and abridged DRM measures were administered to the same sample of participants, the dates of these assessments were separated by approximately three years. Thus, it is impossible to determine whether the moderate correlations across these two methods (correlations in the range of .35 to .40) were due to important differences in methodology, or to the fact that affective experiences are not especially stable across two periods separated by three years. It is important to note however, that correlations across two DRM assessments separated by less than one week only correlated between .45 and .50. Future research could use the full and abridged DRM methods to assess affective experience over the course of the same day to help clarify how much of an impact the specific method has on the results that are obtained.

Conclusion

Researchers interested in subjective well-being have made considerable progress in understanding the individual- and societal-level factors that predict and cause high levels of well-being (Diener et al., 2009). However, much of the research that has been used to inform this understanding relies on broad, global judgments of well-being (e.g., National Research Council, 2014). Although these measures have strong psychometric properties (again, see Diener et al., 2009, for a review), they are not without limitations. Thus, researchers have investigated alternative approaches for assessing well-being, and these alternative, experiential methods have been increasingly used in population-based studies like the PSID. Our study contributes to research on well-being by investigating the psychometric properties of these measures. The results show that aggregated measures of emotions are reliable, stable over the short term, and exhibit meaningful associations with theoretically relevant criteria. Thus, these results providing promising evidence about the utility of experiential measures for the investigation of subjective well-being.

Acknowledgments

References

Anusic I, Lucas RE, & Donnellan MB (2017). The Validity of the Day Reconstruction Method in the German Socio-economic Panel Study. Social Indicators Research, 130(1), 213–232. doi:10.1007/s11205-015-1172-6 [PubMed: 30532344]

Campbell DT, & Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56(2), 81–105. doi:http://dx.doi.org.proxy2.cl.msu.edu/10.1037/h0046016 [PubMed: 13634291]

Carr D, Cornman JC, & Freedman VA (2017). Disability and Activity-related Emotion in Later Life: Are Effects Buffered by Intimate Relationship Support and Strain? Journal of Health and Social Behavior, 58(3), 387–403. doi:10.1177/0022146517713551 [PubMed: 29164961]

Carstensen LL, Fung HH, & Charles ST (2003). Socioemotional Selectivity Theory and the Regulation of Emotion in the Second Half of Life. Motivation and Emotion, 27(2), 103–123. doi:10.1023/A:1024569803230

Cheung F, & Lucas RE (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. Quality of Life Research, 23(10), 2809–2818. [PubMed: 24890827]

Diener E, Emmons RA, Larsen RJ, & Griffin S (1985). The Satisfaction With Life Scale. Journal of Personality Assessment, 49, 71–75. [PubMed: 16367493]

Diener E, Lucas RE, Schimmack U, & Helliwell J (2009). Well-being for public policy. Oxford University Press, USA.

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

Donnellan MB, & Lucas RE (2008). Age Differences in the Big Five Across the Life Span: Evidence From Two National Samples. Psychology and Aging, 23(3), 558–566. [PubMed: 18808245]

Freedman VA, Cornman JC, & Carr D (2014). Is Spousal Caregiving Associated With Enhanced Well-Being? New Evidence From the Panel Study of Income Dynamics. The Journals of Gerontology: Series B, 69(6), 861–869. doi:10.1093/geronb/gbu004

Freedman VA, Stafford F, Conrad F, Schwarz N, & Cornman JC (2012). Assessing Time-Diary Quality for Older Couples: An Analysis of the PSID Disability and Use-of-Time Supplement. Annals of Economics and Statistics, (105/106), 271–289. doi:10.2307/23646465 [PubMed: 23560241]

Hudson NW, Lucas RE, & Donnellan MB (2017). Day-To-Day Affect is Surprisingly Stable: A 2-Year Longitudinal Study of Well-Being. Social Psychological and Personality Science, 8(1), 45–54. doi: 10.1177/1948550616662129 [PubMed: 29238453]

Kahneman D (1999). Objective happiness In Kahneman D, Diener E, & Schwarz N (Eds.), Well-being: The foundations of hedonic psychology (pp. 3–25). New York, NY: Russell Sage Foundation.

Kahneman D, & Krueger AB (2006). Developments in the measurement of subjective well-being. The Journal of Economic Perspectives, 20(1), 3–24.

Kahneman D, Krueger AB, Schkade DA, Schwarz N, & Stone AA (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. Science, 306(5702), 1776–1780. doi:10.1126/science.1103572 [PubMed: 15576620]

Krueger AB, Kahneman D, Schkade DA, Schwarz N, Stone AA, & Krueger AB (2009). National Time Accounting: The Currency of Life In Measuring the subjective well-being of nations: National accounts of time use and well-being (pp. 9–86). Chicago: University of Chicago Press.

Lee Y, Hofferth SL, Flood SM, & Fisher K (2016). Reliability, Validity, and Variability of the Subjective Well-Being Questions in the 2010 American Time Use Survey. Social Indicators Research; Dordrecht, 126(3), 1355–1373. doi:http://dx.doi.org.proxy2.cl.msu.edu/10.1007/s11205-015-0923-8

Lucas RE, Diener E, & Larsen RJ (2003). Measuring positive emotions In Lopez SJ & Snyder CR (Eds.), Positive psychological assessment: A handbook of models and measures (pp. 201–218). Washington, DC, US: American Psychological Association.

National Research Council. (2014). Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience. National Academies Press.

Robinson MD, & Clore GL (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. Psychological Bulletin, 128(6), 934–960. [PubMed: 12405138]

Schimmack U (2008). The structure of subjective well-being In Eid M & Larsen RJ (Eds.), The science of subjective well-being (pp. 97–123). New York: Guilford Press.

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). Russell Sage Foundation.

Smith J, Ryan LH, Queen TL, Becker S, & Gonzalez R (2014). Snapshots of Mixtures of Affective Experiences in a Day: Findings from the Health and Retirement Study. Journal of Population Ageing, 7(1), 55–79. doi:10.1007/s12062-014-9093-8 [PubMed: 24729799]

Stone AA, Shiffman SS, & DeVries MW (1999). Ecological momentary assessment In Kahneman D, Diener E, & Schwarz N (Eds.), Well-being: The foundations of hedonic psychology (pp. 26–39). New York, NY: Russell Sage Foundation.

Watson D, Clark LA, & Tellegen A (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063. [PubMed: 3397865]
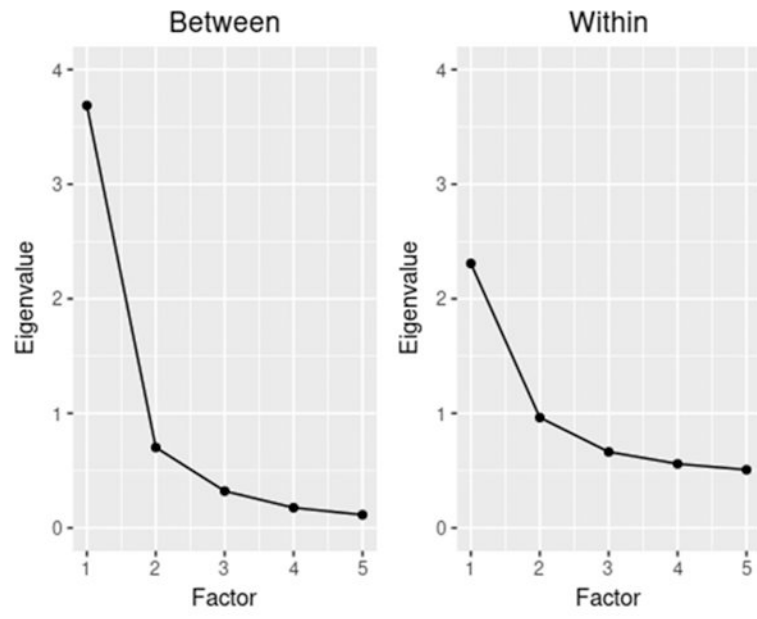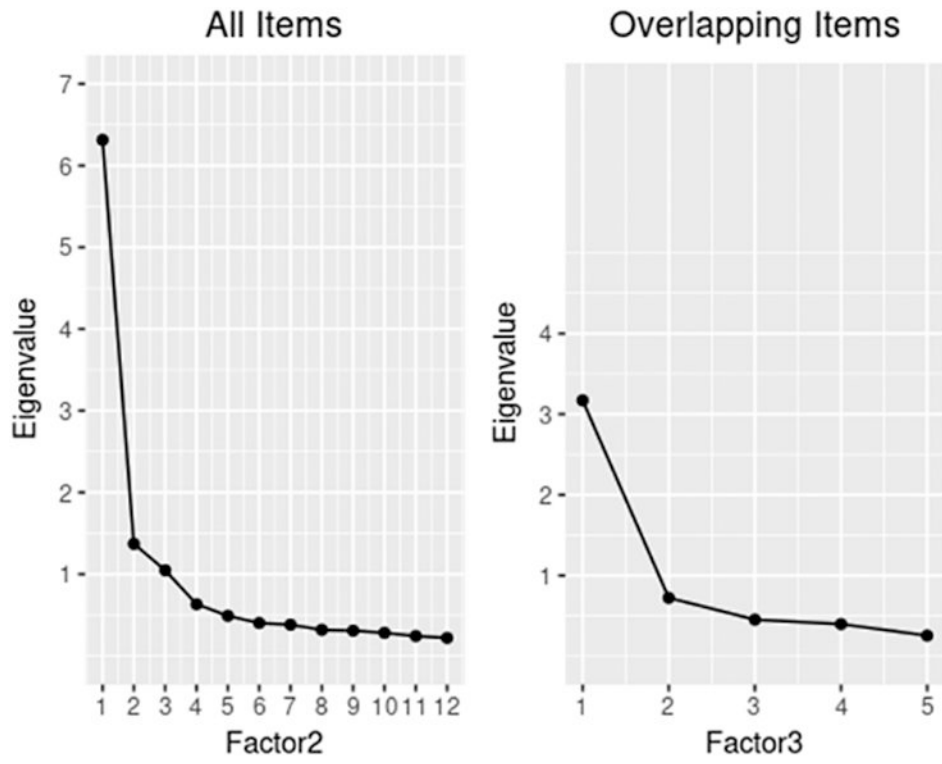
Figure 1.

Figure 2.

Table 1.

Descriptive Data for Emotion Items for the Full and Abridged DRM Measures

	Activity (0–6)		Person (0–6)		Abridged (1–5)	
	mean	sd	mean	sd	mean	sd
Calm	5.31	1.22	5.30	0.75	4.07	0.85
Happy	5.01	1.23	5.02	0.87	3.97	0.94
Frustrated	0.90	1.60	0.90	1.00	1.58	0.83
Worried	0.61	1.34	0.61	0.94	1.59	0.91
Sad	0.46	1.22	0.47	0.89	1.40	0.79
Enthusiastic					3.43	1.15
Content					3.88	1.01
Interested					3.78	1.02
Angry					1.29	0.64
Stressed					1.63	0.86
Lonely					1.39	0.83
Bored					1.45	0.83

Table 2.

Correlations Among Emotion Items from the DUST DRM

	Calm	**Happy**	**Frustrated**	**Worried**	**Sad**
Calm	1.00	0.44	−0.33	−0.24	−0.19
Happy	0.71	1.00	−0.35	−0.28	−0.28
Frustrated	−0.73	−0.60	1.00	0.40	0.30
Worried	−0.64	−0.53	0.84	1.00	0.45
Sad	−0.53	−0.50	0.75	0.85	1.00

Note. Correlations above the diagonal reflect within-person correlations; correlations below the diagonal reflect between-person correlations

Table 3.

Factor Loadings for Factor Analysis of 12 Wellbeing Module Emotion Items

	Factor 1	Factor 2	Factor 3
1. Calm	0.617	−0.428	0.195
2. Happy	0.718	−0.247	0.004
3. Enthus.	0.786	−0.002	−0.081
4. Content	0.721	−0.206	−0.027
5. Inter.	0.741	0.021	−0.153
6. Angry	0.057	0.662	0.001
7. Frustr.	0.026	0.777	0.037
8. Sad	0.000	0.578	0.331
9. Stressed	−0.020	0.797	0.006
10. Lonely	0.000	0.143	0.735
11. Worried	−0.018	0.667	0.134
12. Bored	−0.048	0.021	0.685

Table 4.

Correlations between weekday and weekend reports from the Full-DRM scale scores.

	WD WB	WD PA	WD NA	WE WB	WE PA	WE NA
WD WB	1.00					
WD PA	0.81	1.00				
WD NA	−0.94	−0.57	1.00			
WE WB	0.52	0.43	−0.48	1.00		
WE PA	0.41	0.45	−0.32	0.81	1.00	
WE NA	−0.49	−0.34	0.49	−.94	−0.55	1.00

Note. N = 1693. WD = Weekday, WE = Weekend, WB = mean of all five emotion items, PA = mean of two positive affect items, NA = mean of three negative affect items.

Table 5.

Correlations between full and abridged DRM scale scores.

	Wellbeing	Positive Affect	Negative Affect
Wellbeing (Short)	0.41	0.36	−0.38
Positive Affect (Short)	0.38	0.38	−0.32
Negative Affect (Short)	−0.36	−0.28	0.36
Wellbeing (Long)	0.41	0.37	−0.38
Positive Affect (Long)	0.38	0.38	−0.32
Negative Affect (Long)	−0.37	−0.29	0.36
Boredom/Loneliness	−0.25	−0.19	0.26

Note. N = 1250. Columns are scale scores from the full DRM; Rows are scale scores from the abridged method. Short scales use the five that overlap with the full DRM; Long scales include seven additional items that were only included in the abridged method.

Table 6.

Correlation between DRM-based measures and life satisfaction.

	WB	PA	NA	WB	PA	NA
PSID LS 2009	0.25	0.26	−0.21	0.28	0.29	−0.21
PSID LS 2011	0.31	0.25	−0.30	0.27	0.27	−0.23
PSID LS 2013	0.31	0.28	−0.28	0.32	0.35	−0.23
DUST LS 2013a	0.35	0.37	−0.28	0.34	0.36	−0.27
DUST LS 2013b	0.39	0.41	−0.32	0.34	0.35	−.27

Note. WB = Wellbeing, PA = Positive Affect, NA = Negative Affect, LS = Life Satisfaction.

Table 7.

Correlation between DRM-based measures and personality and spirituality.

	WB	PA	NA	WB	PA	NA
Extraversion	0.09	0.09	−0.08	0.13	0.14	−0.10
Agreeableness	0.13	0.14	−0.11	0.07	0.08	−0.05
Conscientiousness	0.17	0.12	−0.17	0.13	0.13	−0.11
Neuroticism	−0.34	−0.29	0.32	−0.34	−0.33	0.30
Openness	0.00	−0.04	−0.02	0.05	0.07	−0.03
Self-efficacy	0.26	0.21	−0.26	0.26	0.27	−0.21
Spirituality	0.10	0.17	−0.04	0.07	0.08	−0.04

Note. WB = Wellbeing, PA = Positive Affect, NA = Negative Affect.

Table 8.

Correlation between DRM-based measures and Relationship Variables.

	WB	PA	NA	WB	PA	NA
Relationship Quality	0.29	0.29	−0.25	0.23	0.25	−0.19
Family Quality	0.16	0.16	−0.15	0.19	0.15	−0.18
Number of Friends	0.12	0.13	−0.10	0.12	0.13	−0.09

Note. WB = Wellbeing, PA = Positive Affect, NA = Negative Affect.

Table 9.

Correlation between DRM-based measures and Measures of Disability.

	WB	PA	NA	WB	PA	NA
Impairment	−0.27	−0.22	0.26	−0.24	−0.23	0.22
Activity Limitations	−0.20	−0.13	0.21	−0.18	−0.17	0.17
Memory Rating	−0.19	−0.15	0.18	−0.19	−0.18	0.14
Memory Change	−0.12	−0.13	0.10	−0.11	−0.13	0.09
Memory Aids	−0.14	−0.16	0.10	−0.11	−0.09	0.11

Note. WB = Wellbeing, PA = Positive Affect, NA = Negative Affect.