HHS Public Access

Author manuscript

Soc Indic Res. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

Soc Indic Res. 2016 April 1; 126(3): 1355–1373. doi:10.1007/s11205-015-0923-8.

Reliability, Validity, and Variability of the Subjective Well-Being Questions in the 2010 American Time Use Survey

Yoonjoo Lee,

Sandra L. Hofferth,

Sarah M. Flood, and

Kimberly Fisher

Abstract

Part of a wider range of investigations to produce generally acceptable standards for measuring affective well-being, time diary surveys have tested several approaches to measuring subjective well-being during diary days. As an alternative to the standard approach of asking a single question about each activity reported in time diary surveys, the 2010 module of the American Time Use Survey asked six emotion questions about three activities. The perception questions captured how happy, meaningful, sad, tired, stressed, or in pain respondents felt on a 7-point scale. To evaluate this approach, our research examined the reliability and validity of the six emotion questions, and assessed their variability across activities. Using principal component analysis, we assessed the associations among items and obtained two activity-level components with Cronbach's alphas of 0.68 and 0.59 and two respondent-level components with Cronbach's alphas of 0.74 and 0.65. To test validity, we regressed self-rated health on the underlying components and socio-demographic controls. Both of the respondent level components were significantly associated with better health (odds ratio 1.81, 1.27). Using each of the perceptions individually, we found that happiness, meaningfulness, and lack of fatigue, stress, and pain were related to better health, but none as strongly as the first component. Finally, we examined the coefficients of variation to assess the variability in the well-being measures across activities. Measurement implications and limitations of this study are discussed.

Keywords

Subjective well-being; Time use; Health; Methodology

1 Introduction

Although there is wide recognition that financial measures are not the only indicators of a good society, the general well-being of a society is usually measured in terms of GDP, national income and consumption (Krueger et al. 2009b). To complement these financial measures, researchers and policymakers have made considerable progress in incorporating

ylee228@umd.edu.

subjective or affective measures, such as trust or happiness, in the discourse regarding the well-being of society. The current study contributes by examining one recent assessment of well-being, a module of the American Time Use Survey (ATUS) that asks individuals to evaluate and report their subjective experiences during three randomly selected activities. Because there are few previous analyses to model and because the specific way activities are sampled makes their analysis complex, this paper provides a descriptive examination of these measures. The objective of this paper is to evaluate the reliability of the measures, their validity in terms of predicting self-rated health, and the variability of self-rated subjective evaluations across different activities.

1.1 What is Subjective Well-Being?

Krueger and colleagues have argued that, ideally, one would like to know the proportion of daily time that an individual experiences negative/positive emotions (Krueger et al. 2009b). The higher the proportion of time positive experience is reported, the higher the well-being. This would be possible only if the investigator were to measure subjective experiences continuously throughout an entire day. Lacking a comprehensive measure, well-being is defined in terms of individual experiences either at specific times or during participation in specific activities.

1.2 Assessment of Subjective Well-Being

Initially, measures of subjective well-being in time diary surveys asked respondents' satisfaction with or enjoyment of daily activities in general. The 1965–1966 time use survey was the first to collect affect information about activities in the U.S. After providing a 24-h recall of the previous day's activities, respondents reported their most enjoyable and least enjoyable activity during their diary day (Robinson and Martin 2009). Similarly, in the second and fourth waves of the 1975 Time Use in Economic and Social Accounts study, collected by Juster and colleagues at the University of Michigan, respondents rated 22 activities on a 10-point scale (from 'dislike a great deal' to 'enjoy a great deal') (Juster et al. 1975). These general ratings are limited in that they may not represent how respondents actually experienced their activities (Krueger et al. 2009a).

Since the 1980s, time diary survey respondents have rated subjective well-being during specific activities. A limited number of time use surveys conducted in Canada, the U.S., and the U.K. experimented with asking a single emotion question of all activities. In the 1985 study of American's Use of Time, individuals rated enjoyment of each reported activity on a 1–10 scale similar to that used in the 1975 survey (Robinson 2013). The 1986 Unilever Household Research Project in the U.K. also asked an enjoyment question for each activity reported in the diary (Gershuny 2013). These approaches yielded an individual's rating of a specific activity episode. In the Family Time and Activity study of Canadians, Michelson (1985) asked the degree of tension experienced in each of the episodes on a 7-point scale (from 'very tense' to 'very relaxed'). Limited published research made use of the affect dimension of these surveys until recent years. While these surveys remained accessible, if little used, two other approaches to collecting emotions alongside daily activities also developed in the later years of the last century.

The Experience Sampling Method (ESM) collects momentary perceptions of subjective states within minutes of the event, as opposed to later on the same day or on the next day (Hektner et al. 2007). Individuals are contacted at random intervals during the day and asked a set of questions about what they are doing and how they feel about it. Studies using the ESM have tended to focus on small local areas or subgroups instead of national populations. Because of its random selection of intervals, the ESM does not obtain information about the full day but can use the sequence of activities across a day to examine the quality of feelings (Kubey and Csikszentmihalyi 1990).

The Princeton Affect and Time Survey (PATS) and earlier smaller scale surveys collected by Kahneman and colleagues included the same fields present in conventional time diary surveys [though Kahneman and Krueger describe their surveys as a Day Reconstruction Method (DRM)], but claimed they added an innovation not present in previous time use surveys (Kahneman et al. 2004). After the day reconstruction was completed, respondents were asked to report on six emotions (pain, happy, sad, tired, stressed, and interested) they may have experienced during three randomly selected 15-min intervals over the course of the day. A limitation of this strategy is that the selected time intervals could have included multiple activities during which individuals might, in some circumstances, experience different emotions. The use of the DRM to collect affect information has been recently questioned (Diener and Tay 2014) due to the paucity of documentation on the scale employed and the six questions chosen (but see National Research Council 2012).

The ATUS well-being module broadly adopted the DRM format in the PATS, but focuses on the activity as opposed to a randomly selected time interval. It asks about meaningfulness rather than interestedness as well as whether the respondent was in pain, happy, sad, tired, and stressed. Each of the six measures is coded ‘not at all’ (0) to ‘very’ (6) for three randomly selected activities that lasted at least 5 min, excluding sleeping, grooming, and personal activities.

With the general agreement that affect information is important for developing national well-being accounts, agencies such as the OECD (2013) and the United Nations Economic Commission for Europe (UNECE 2013) recommended the collection of at least some form of affect data in national surveys. The 2009–2010 French and 2014–2015 British national time use surveys conducted following the Harmonised European Time Use Survey guidelines asked a sub-sample of participants to complete a single enjoyment question for all activities. As a part of the on-going debate towards developing an internationally accepted standard, as recommended by the UNECE (2013), this paper undertakes an in-depth investigation of the method that was chosen for the 2010 ATUS.

1.3 What are the Assumptions Underlying the Measurement of Well-Being?

Based on their research with the PATS, Krueger et al. (2009b) recommended the construction of an index called the “U index” which is created by classifying an activity episode as unpleasant if the most intense feeling reported for that activity is negative. This assumes that the relative ordering on a scale of assessing pain is comparable to the relative ranking on a happiness scale, for example, which may or may not be the case. No empirical work was conducted to establish equivalence. In addition, the result of the U index is a

dichotomous measure—unpleasant or not unpleasant. Many activities have some elements of both pleasantness and unpleasantness. Recent experimentation using emotion measures attached to the Disability and Use of Time supplement of the U.S. Panel Survey of Income Dynamics (Freedman et al. 2014) suggests that some emotions do not change significantly with duration of activities, with the exception of frustration, and that many measures of negative emotion (sad, tired, pain, worried, and frustrated) are strongly correlated. In pursuing a more nuanced measure, we challenge the assumption that these six elements represent a single valid and reliable scale. To do so, we need to understand how they work together.

In this paper we take a more conventional approach, in which we measure the extent of positive reaction to an activity, treated as a continuous construct. We use standard measurement theory, which assumes that there is a dimension of satisfaction or positiveness underlying the experience of each activity episode. The true measure is a function of our observed indicators plus error. Of an infinite set of indicators, the scale measures only a few. Some are worded in a negative and some in a positive direction, but, if they tap a similar dimension, they will be highly correlated. By using standard psychological measurement tools, we can gain a better understanding of the underlying correlation of each item with the underlying construct and, if needed, divide the indicators into more homogeneous groups based on correlations with the underlying construct.

Specifically, in our analysis we focus on reliability, validity, and variability. We evaluate *reliability* using Cronbach's alpha, a measure of the dependability or consistency of the measure: the extent to which the observed items are linked to the true underlying scores.

Validity, in contrast, tells us whether our underlying measure is linked to an hypothesized outcome, in this case self-reported health. Does it measure what it is supposed to measure? Research has consistently shown that happiness and health are correlated (de Mello and Tiongson 2009; Huppert 2009; Veenhoven 2006). We argue that individuals with better subjective well-being should also report better health, without implying any causal linkage.

Finally, although we may be able to adequately measure subjective emotional states during activities, and these may link to hypothesized outcomes, how much variability is there in emotional states across activities and across individuals? Are there differences in emotional states across activities, or do these subjective emotional states appear to better describe the individuals themselves rather than activities? Are these truly momentary perceptions or do they represent consistent underlying individual characteristics?

2 Methods

2.1 Data

The ATUS is a nationally representative annual survey of the time use of Americans aged 15 or older conducted by the U.S. Bureau of Labor Statistics from 2003 to the present (Bureau of Labor Statistics 2014). The monthly Current Population Survey (CPS) provides the sampling frame for this survey; households that complete the eighth and final CPS interview become eligible for selection into the ATUS sample. Because households in small states are

oversampled in the CPS, the first procedure is to eliminate this overrepresentation (Abraham et al. 2006). The second procedure is to stratify households based on the race/ethnicity of the householder, the presence and age of children, and the number of adults in adult-only households, and the sampling rates vary depending on these strata. One person aged 15 or older is randomly selected from each sampled household for the ATUS interview, which occurs 2–5 months after the completion of the last CPS interview.

Once the selected participant agrees to participate in the ATUS interview, the survey is collected by telephone, either in English or Spanish. The ATUS participant is asked to answer questions about socio-demographic characteristics and to report retrospectively how he or she spent one 24-h day. The respondent describes his or her primary activities, with whom he or she was at those times, and the place where the activities took place. The average response rate from 2003 to 2013 was 55 % (ranging from 57.8 % in 2003 to 49.9 % in 2013). The response rate has not seriously affected the generalizability of the survey, but those who are weakly involved in their communities have been shown to be less likely to respond (Abraham et al. 2006). ATUS interviews are conducted every day including holidays (New Year's Day, Easter, Memorial Day, 4th of July, Thanksgiving, and Christmas); no interviews were conducted on Thanksgiving or Christmas Day in 2003 and on Thanksgiving Day since 2004. Ten percent of the ATUS diary days are assigned to each of the weekdays, 25 % are assigned to Saturdays, and the remaining 25 % to Sundays, and they are also distributed evenly across the weeks of the year. Using population weights, the data are adjusted to represent actual weekdays and weekend days in a year.

The data for this study were drawn from the 2010 well-being module of the ATUS. We make use of the version of the ATUS available through the ATUS-X Extract Builder system (Hofferth et al. 2013; http://www.atusdata.org), which offers the advantage of allowing simplified linking of supplementary data (including the emotion supplement) and variables from previous waves of the CPS to the ATUS data. The 2010 well-being module is the first ATUS supplementary module to collect information on the respondent's subjective well-being during three randomly selected activities, each of which lasted at least 5 min.

Because the respondent is also asked to report self-rated health status, the 2010 well-being module can be used to examine diverse aspects of the respondent's well-being, in combination with his or her time use patterns. We focused on ATUS well-being module respondents aged 25–64 because those who are younger or older may have different health conditions or disabilities and different activities compared with those in the prime employment years. Also, the pressures of work-life balance fall most heavily on people of working age (Drobnic and Guillén 2011); hence, this group is of high policy significance when measuring national emotional well-being. We limited our sample to those with complete socio-demographic information, time diaries, and at least one subjective well-being report for one of the randomly selected activities. Given these conditions, our sample consists of 8791 respondents to the ATUS well-being module. Descriptive information for the sample is presented in “Appendix” Table 8.

We created two analysis files: one for activity-level data and one for respondent-level data. An activity-level file, called an event or episode file, has one record (row) for each activity,

along with variables identifying the type of activity, its start and end time, the duration of the activity, and its location (Michelson 2005). Because the typical respondent engages in many activities during a diary day, one respondent has multiple records of activities and the number of records varies depending on the respondent. The activity-level ATUS data in this paper consist of the respondent's assessed activities and his or her six subjective well-being reports for up to three activities, with everyone reporting at least one activity: 97.52 % have three complete activities and emotion reports (8573 respondents × 3 activities = 25,719 records); 2.26 % have two complete activities and emotion reports (199 respondents × 2 activities = 398 records); and 0.22 % have only one activity and its emotion report (19 respondents × 1 activity = 19 records). Including those with one, two, or three selected activities, the total number of records in the activity-level data is 26,136 (=25,719 + 398 + 19).

A respondent-level file, called a summary file, has one record (row) for each respondent, along with the columns including aggregated values of time-use components (Michelson 2005). In the ATUS, for instance, 20 min of walking in the morning and 30 min of walking in the evening are coded as two separate records in the activity-level data. In the respondent-level file, however, the respondent is reported as walking for 50 min on the diary day. The respondent-level file can be used to calculate summary measures of time use for populations or subgroups and to examine individual characteristics that may be associated with them. The respondent-level well-being data in this paper come from averaging emotion scores across (up to three) records in the activity-level file for each respondent; the total number of records in the respondent-level data file is 8791 (=8573 + 199 + 19).

2.2 Measures

The subjective well-being measures used in this study comprised two positive (happy and meaningful) and four negative emotions (sad, tired, stressed, or in pain). Five of the six emotions were asked about each of the randomly selected activities in a similar format: “From 0 to 6, where a 0 means you were not happy/sad etc. at all and a 6 means you were very happy/sad etc., how happy/sad etc. did you feel during this time?” The sixth measure was asked as follows: “From 0 to 6 where a 0 means not meaningful at all to you and a 6 means very meaningful to you, how meaningful did you consider what you were doing?” The order of the six questions was randomly determined for each respondent.

The self-rated health question was used to validate the subjective well-being measures. The ATUS participant was asked, “Would you say your health in general is excellent, very good, good, fair, or poor?,” on a scale of 1 (excellent) to 5 (poor).

In examining the association between subjective well-being and self-rated health, we adjusted for socio-demographic factors that may affect well-being and may also affect health (Haas 2007; Huppert 2009), including age, educational attainment, the family's total annual income, employment status, race/ethnicity, gender, marital status, whether the respondent had a child under the age of 18 in the household, whether the respondent had any physical or cognitive difficulty, and the amount of time spent sleeping.

2.3 Analysis Plan

To assess the reliability of the six subjective well-being measures, we conducted a principal component analysis and calculated Cronbach's alphas. Because the four negative emotions were asked on the same scale as the two positive emotions, either the four negative or the two positive emotions could be recoded for consistency and easier interpretation. We reverse-coded the four negative emotion items so that each item was an indicator of positive subjective well-being. All of the statistics for the reliability test were calculated both for activity-level and for respondent-level well-being measures.

Validity was tested by regressing self-rated health on the subjective well-being indicators, using an ordinal logistic regression. This health question has five ordered categories, which makes the use of ordinal logistic regression appropriate. Health is coded so that a positive association means that a variable is associated with improved health on this scale. The socio-demographic controls were included in the validity assessment model, thereby obtaining a clearer picture of whether perceived well-being was independently associated with general health status. Because the self-reported general health indicator and socio-demographic variables were measured at the respondent level, the validity test could not be conducted at the activity level. To show how the six items can be used in different ways in the validity test, three types of well-being measures (scores obtained from a single factor and from a two-factor principal component analysis and individual items) were utilized in the analysis models.

The variability of subjective emotion evaluations across activities was tested by the coefficient of variation (CV), the ratio of the standard deviation to the mean. Because the CV indicates the extent to which an item of interest varies, the higher the CV, the greater the variability of the six perceptions across activities and across people. The CVs were calculated at both the activity level and at the respondent level, and all of these analyses were conducted using SAS 9.3. Analyses were weighted using the final activity-level weight for activity-level data and the final respondent-level weight for respondent-level data.[1]

3 Results

3.1 Descriptive Statistics

Means and standard deviations of the six subjective well-being measures are shown in Table 1. The activity-level means are the average emotion scores of all the activities; the respondent-level means are the averages of individuals' emotion scores, which have been averaged across the assessed activities. Activity-level and respondent-level mean scores of the same emotional state were very similar. No sadness had the highest mean in both of the data sets, followed by no pain, indicating that respondents reported very little sadness and pain, on average. These results confirm the finding in previous studies conducted with an early version of the DRM that negative emotions were rarely observed (Kahneman and Krueger 2006). Fatigue showed a relatively higher mean score than the other negative

[1]The BLS reported a problem with the selection of activities other than sleeping in the last episode for the 2010 and 2012 ATUS module and re-released corrected weights on May 21, 2014. This study used the 2010 well-being module data containing the weights adjusted for this error. For more information see http://www.bls.gov/tus/wbnotice.htm; accessed July 9, 2014.

subjective well-being measures. For the two positive feelings, the mean of meaningfulness was higher than that of happiness in the two data sets, but the two scores in the respondent-level data did not differ.

All of the six emotional states were significantly different from each other in the activity-level data [$F(5, 156{,}635) = 2945.40$, $p < .001$]. The same pattern was revealed in the respondent-level data [$F(5, 52{,}740) = 1417.45$, $p < .001$], except for the relation between the two positive feelings—meaningfulness and happiness. The differences in the subjective well-being measures will be further discussed in Sect. 3.4.

3.2 Reliability

Table 2 presents correlations among the subjective well-being measures. Overall, the two positive emotions were moderately correlated with one another; the four negative feelings were also closely related. The correlations between positive and negative items were relatively low. For instance, the meaningfulness item was moderately correlated with happiness (0.40 in the activity-level data; 0.48 in the respondent-level data), but had very weak correlations with the negative emotional states (0.00–0.06 in the activity-level data; 0.01–0.08 in the respondent-level data). Although the two adjectives of ‘happy’ and ‘sad’ are antonyms from a linguistic perspective and paired as polar opposites (Osgood et al. 1957), happiness and no sadness showed relatively low correlations (0.32 in the activity-level data; 0.37 in the respondent-level data).

Using the six emotion items, principal component analysis with varimax rotation was conducted to identify the components in both the activity-level and respondent-level data sets. We first requested only one component (Table 3, upper panel). The six items explain 39.21 % of the variance in the variables in the activity-level and 43.42 % of the variance of the variables in the respondent-level data. We then used two criteria to determine the number of components extracted: eigenvalue and scree plot. An eigenvalue is defined as “the amount of variance that is captured by a given component” (O'Rourke et al. 2005), and any component with an eigenvalue of 1 or greater is considered appropriate to retain. A scree plot graphically shows the eigenvalue of each of the components, and any component before a noticeable drop is assumed to be retained. These two criteria in our analysis suggested two components in each data file (Table 3, lower panel). Component 1 in the activity-level data had an eigenvalue of 2.35 and that in the respondent-level data had an eigenvalue of 2.61. Each component accounted for the same amount of variance as the sole component in the aforementioned one factor analysis in the respective data files. Component 2 in the activity-level data had an eigenvalue of 1.25 and explained 20.87 % of the variance; that in the respondent-level data had an eigenvalue of 1.31 and explained 21.79 % of the variance. The grouping of the items in the principal component analysis is similar to the prior analysis (Table 2) because it reflects the underlying correlations. However, it is possible that the components are mainly driven by the question wording (positive or negative direction) and devoid of substantive meaning because happiness and no sadness were not grouped together. To address this we also conducted principal component analysis using only five items (happiness, no pain, no sadness, no stress, and no fatigue), given the low correlations between meaningfulness and other items (Table 3, upper panel). Only one

component was suggested with eigenvalues of 2.31 in the activity-level data and of 2.56 in the respondent-level data. This suggests that the second component results from the inclusion of meaningfulness and not from the positive–negative wording of the items.

Internal consistency coefficients (Cronbach's alphas) were also calculated for both data sets. When we requested one factor from all six items, its Cronbach's alpha was 0.66 in the activity-level data and 0.70 in the respondent-level data. When we obtained two factors, the four items in the first component showed a Cronbach's alpha of 0.68; the two items in the second component had a Cronbach's alpha of 0.59 in the activity-level data. In the respondent-level data, we obtained Cronbach's alphas of 0.74 from the four items in the first component and 0.65 from the two items in the second component. The acceptable range for Cronbach's alpha varies by number of items, variability, and purpose of the scale. There is no fixed cutoff but a range of 0.60–0.80 is acceptable (Nunnally 1978). Therefore, the six items in the 2010 well-being module of the ATUS showed acceptable Cronbach's alphas in the activity-level and respondent-level data sets.

3.3 Validity

We tested the validity of the six averaged respondent-level subjective well-being items by examining their association with the self-rated general health indicator. This convergent validity was assessed in three models in Table 4. Model 1 shows how the underlying constructs found in the principal component analysis were linked to general health status. Factor 1 (no pain, no sadness, no stress, and no fatigue) was significantly associated with better health [Spearman Correlation 0.32; odds ratio (OR) 1.81]. Factor 2 (happiness and meaningfulness) was also significantly linked to better health (Spearman Correlation 0.07; OR 1.27). Model 2 presents the result of regressing general health on each of the subjective well-being items and socio-demographic controls. Although happiness (Spearman Correlation 0.18; OR 1.18), meaningfulness (Spearman Correlation 0.04; OR 1.06), lack of pain (Spearman Correlation 0.35; OR 1.42), stress (Spearman Correlation 0.20; OR 1.06), and fatigue (Spearman Correlation 0.21; OR 1.12) were significantly related to better health, lack of sadness was not. We used the single factor obtained earlier in Model 3, and this measure was also significantly associated with better health (Spearman Correlation 0.33; OR 1.89). Using only five items (without meaningfulness), instead of six, produced an OR of 1.89 and a Spearman's correlation of 0.33, $p < .001$. Our findings indicate that individuals with better subjective well-being also reported better health.

Given the various measures reviewed, in Table 5 we summarize the reliability and validity tests and potential tradeoffs across four potential combinations of the six well-being measures.

3.4 Variability

The CVs across subjective well-being items at both the activity and respondent levels, calculated by dividing the standard deviations by the means presented in Table 1, indicate the amount of variability (Table 6). Because the no fatigue item showed the highest standard deviation and the lowest mean, its CV was the highest. The lowest CV was found in the no sadness item due to its low standard deviation and high mean. Additionally, the CVs in the

activity-level data were greater than those in the respondent-level data. This makes sense because the emotion scores in the respondent-level data are obtained by averaging the average of each respondent's three well-being assessments. The variability decreases when the emotion measures are averaged across activities.

We illustrate the variability of the six affect items using box plots for the activity-level data in Fig. 1 and the respondent-level data in Fig. 2. The three lines in each box indicate the first, second (median), and third quartiles of a variable. Thus, the box represents how the data from the first quartile to the third quartile were distributed. Because the no pain item and the no sadness item showed highly skewed distributions, their second quartiles were the same as their third quartiles, and their boxes show only two lines. The different sizes of the emotion item boxes correspond to the differences in the CVs across measures in both data sets presented in Table 6. Given the CVs and box plots in the table and figures, our findings indicate relatively small to moderate variability across different subjective well-being measures.

Table 7 presents correlations and weighted means and standard deviations of the emotional states across activities. The means were calculated by dividing the sum of the positive perceptions by the number of selected activities, so they represent how positively the respondents assessed the assigned activities. The means of the first and third activities were around 4.50, and the mean of the second activity was slightly lower, 4.45. To assess the variability of the emotional states across activities, mean differences were tested using the GLM test. Means across activities were statistically significant [$F(2, 26{,}133) = 7.14$, $p < .001$]; differences were detected between the first and second activities and between the second and third activities at the $p < .05$ levels. No matter what types of activities the respondent assessed, he or she reported significantly lower positiveness in the second activity than in the first or third activities.

Emotional states across activities showed relatively strong correlations. Adjacent activities were correlated around 0.70, but the first and third activities were correlated at 0.63. The Cronbach's alpha of the three assessments is 0.86, which suggests high consistency across the respondents' reports. The CVs across the respondent's three activities are similar at about 23 %, indicating low variability. Thus, the small to moderate variability across the subjective well-being measures as well as across the three activities and the high correlations across activities suggest that the respondent's emotional reports on the three randomly selected activities are closely related to individual-specific characteristics, such as positiveness. Consistent with this result, Patulny and Fisher (2012) found that the context as well as activities themselves were associated with affective responses.

4 Discussion

The ATUS 2010 well-being module represents one methodology for combining subjective well-being reports during a randomly selected set of activities with time diaries. Given that the subjective well-being measure therein represents an individual's positive or negative experiences, the six affect items (happiness, meaningfulness, in pain, sadness, stress, and fatigue) can be utilized to study the ATUS respondents' well-being in diverse ways. Using

the average scores of the affect items in the activity-level data and in the respondent-level data, we examined the reliability, validity, and variability of the measures with a sample of 8791 ATUS well-being module respondents aged 25–64.

4.1 Use of All Six Items in One Scale

If the researcher wishes to use all six items in one scale, our reliability tests of the six emotion items show sufficiently high reliability estimates. The Cronbach's alpha in the activity-level data was 0.66 and that in the respondent-level was 0.70. They are in line with reliability estimates in other subjective well-being measurement approaches, such as the ESM or the DRM. In the ESM, for example, the internal consistency reliability among four affect measures (happy, cheerful, sociable, and friendly) was 0.57 (Csikszentmihalyi and Larson 1987; Csikszentmihalyi and Hunter 2003). The test–retest reliability scores of the subjective well-being measures in the PATS DRM were 0.62 (happiness) and 0.60 (depressed) (Krueger and Schkade 2008). The weak item is meaningfulness, with a low correlation between meaningfulness and other items and a weak linkage between the meaningful item and a single underlying component (0.26 for activity-level data and 0.27 for respondent-level data).

To use all six items in one scale, the analyst would output a weighted factor score for the entire six-item scale; the weights for the meaningfulness item suggested in principal component analysis are 0.11 in the activity-level data and 0.10 in the respondent-level data. Compared to the weights of other items (0.25–0.33 in the activity-level data; 0.24–0.31 in the respondent-level data), these low weights represent very little contribution of the meaningfulness item to the single component. An alternative strategy would be to drop meaningfulness. The five items without meaningfulness fit well into one scale. At the respondent level the five items explain 51 % of the variance compared with 43 % of the variance using all six items.

4.2 Division into Two Scales

Given the low correlations between meaningfulness and other items, creating two scales is a reasonable option. The six emotion items can be grouped into two underlying components. The meaningfulness item showed the biggest increase in factor loadings when we specify two factors instead of one factor in the principal component analysis (0.26–0.86 in the activity-level data, 0.27–0.89 in the respondent-level data) and the proportion of variance explained increased as well, from 43 % of the variance with a single six-item scale to 65 % with two scales. The first component consists of the four items about the lack of negative emotions; the second component has two positive feelings. Recoding the four negative emotion items did not alter their association with the positive emotion indicators. No matter how they are worded, the four negative feelings are more related to each other than to the positive feeling indicators, and they tap a different construct compared with the positive emotions. We infer that the lack of negative feelings is not strongly linked to the existence of positive feelings.

In this paper we used all items and obtained two separate scales from principal component analysis. These two components explain over 60 % of the variance in the activity-level data

and the respondent-level data, and the meaningfulness item and the happiness item show similar weights needed to calculate factor scores (weights in the activity-level data: 0.65 and 0.53, respectively, those in the respondent-level data: 0.64 and 0.52).

The convergent validity test of the six subjective well-being measures shows that individuals with better subjective well-being, such as higher levels of positive emotions and lower levels of negative ones, report better health status, as hypothesized. Both components obtained in the principal component analysis were significantly associated with health.

4.3 Individual Items

We do not recommend using individual items. Attempting to obtain separate independent effects while controlling for correlated items will not produce as high an association with the outcome as will combining related items into separate orthogonal scales. When we include each of the emotion items individually in the model, the item asking about sadness is not as effective as other items in predicting better health because it is highly correlated (0.60 in Table 2) with stress. None of the items individually is as strong a predictor of health as is the first component, and only “no pain” is stronger than the second component of the two-factor model.

Although in this paper we tested the validity of the six items by looking at the association between a health indicator and respondent-level well-being measures, this is not the only alternative. Robinson (2013) took the approach of comparing the average of five of the six items in the 2010 Well-being module in the ATUS to activity enjoyment ratings in the 1975 University of Michigan time-use survey and the 1985 University of Maryland time-use survey in order to see whether the measures in the ATUS replicate previous attempts to capture subjective well-being. He found that the measures in the ATUS largely replicate enjoyment ratings in the previous surveys, another demonstration of the well-being scale's validity. A question asking the respondent's overall life satisfaction was newly introduced in the 2012 Well-being module (Bureau of Labor Statistics 2014). This item can be used as a criterion measure to test the six well-being items’ validity in the future because higher positive well-being scores are likely to predict greater life satisfaction, along with better health.

The variability test shows that there is small to moderate variability across the affect items. Given that the six emotion items are grouped into two underlying constructs, and the types of activities that the respondent does would not affect their emotion reports in this study due to the random selection of the activities, we infer that individual characteristics, such as optimism or positivity, are reflected in the six subjective well-being measures, rather than activity-specific emotions. This is supported by the high correlations and the small variability across the three activities. Additionally, because the emotion report was collected retrospectively after the respondent completed his or her time diary, the exact emotions that the respondent felt while engaging in the randomly selected activities may not have been captured precisely.

Several limitations of this study should be noted. First, the validity of the subjective well-being measures could not be tested with other approaches, such as divergent validity, due to

the lack of appropriate variables in the data. Although the convergent validity test that we used in this study is one of the most common methods to test construct validity (Marczyk et al. 2005) and it showed significant findings, diverse methods to validate the subjective well-being items would be helpful to provide stronger support for the measures (see Lyubomirsky and Lepper 1999; Robinson 2013). Second, the ATUS respondents may differently interpret and assess the randomly selected activities on the same scale, depending on the emotions assessed. For instance, one may assess a selected activity as being 'very' sad, but the intensity of being 'very' sad may not be the same as that of being 'very' meaningful. Similarly, one degree change of sadness may not be equally perceived as that of meaningfulness. We compared the six emotion items measured on the same scale, assuming that the scale of 0–6 is equally understood across the six well-being measures. Third, the selected activities' duration or timing over the diary day may affect respondents' reports on the activities. In their ESM and DRM research, Kahneman and Krueger (2006) found that people felt least tired around noon, and they felt more tired after noon. In further analysis, controlling for the amount of time the respondents spent during the selected activities (ranging from 5 min to 18 h 27 min) did not change the convergent validity test results (not shown), but *when* the activities took place over the diary day could not be taken into account because of our usage of average emotion scores.

Despite their limitations, the affect items in the well-being module of the ATUS show sufficient reliability and validity, and moderate variability, all of which improve their value for research. This paper demonstrates the implications of different uses of the affect data in the publicly available ATUS. Researchers should keep in mind that our suggestions may not be applicable to affect data in time use surveys utilizing different methods for measuring subjective well-being.

Acknowledgments

A preliminary version of this paper was presented at the 35th International Association of Time Use Research (IATUR) Conference in Rio de Janeiro, Brazil, in 2013. Support for this research was provided by the National Institute of Child Health and Human Development (Grant number R01-HD053654, S. Hofferth, PI, and R24-HD041041, the Maryland Population Research Center).

Appendix

See Table 8.

Table 8

Weighted means and standard deviations of socio-demographic variables in the sample

	Mean	SD
General health		
Excellent	0.20	
Very good	0.34	
Good	0.28	
Fair	0.13	

	Mean	SD
Poor	0.04	
Socio-demographic variables		
Age	44.02	10.84
Education in years	13.82	3.03
Family income		
Lowest quintile	0.13	
Second quintile	0.21	
Third quintile	0.18	
Fourth quintile	0.26	
Highest quintile	0.22	
Employment status (1 = working)	0.75	
Race/ethnicity		
White	0.67	
Black	0.12	
Asian	0.04	
Hispanic	0.14	
Other	0.01	
Gender (1 = male)	0.49	
Children < age 18 in the household (1 = yes)	0.44	
Marital status (1 = married)	0.63	
Sleep time (Divided by 10 min)	30.76	9.93
Any physical/cognitive difficulty (1 = no)	0.93	
N	8791	

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

References

Abraham K, Maitland A, Bianchi S. Nonresponse in the American Time Use Survey: Who is missing from the data and how much does it matter? Public Opinion Quarterly. 2006; 70(5):676–703. doi: 10. 1093/poq/nfl037.

Bureau of Labor Statistics. [9 July 2014] American Time Use Survey. 2014. http://www.bls.gov/tus/.

Csikszentmihalyi M, Hunter J. Happiness in everyday life: The uses of experience sampling. Journal of Happiness Studies. 2003; 4(2):185–199.

Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. Journal of Nervous and Mental Disease. 1987; 175(9):526–536. [PubMed: 3655778]

de Mello L, Tiongson E. What is the value of (my and my family's) good health? Kyklos. 2009; 62(4): 594–610. doi:10.1111/j.1467-6435.2009.00453.x.

Diener E, Tay L. Review of the day reconstruction method (DRM). Social Indicators Research. 2014; 116(1):255–267. doi:10.1007/s11205-013-0279-x.

Drobnic, S.; Guillén, AM., editors. Work-life balance in Europe: The role of job quality. Palgrave Macmillan; Basingstoke, UK: 2011.

Freedman V, Conrad F, Cornman J, Schwarz N, Stafford F. Dose time fly when you are having fun? A day reconstruction method analysis. Journal of Happiness Studies. 2014; 15(3):639–655. doi: 10.1007/s10902-013-9440-0. [PubMed: 25110460]

Gershuny J. National utility: Measuring the enjoyment of activities. European Sociological Review. 2013; 29(5):996–1009. doi:10.1093/esr/jcs077.

Haas S. The long-term effects of poor childhood health: An assessment and application of retrospective reports. Demography. 2007; 44(1):113–135. [PubMed: 17461339]

Hektner, J.; Schmid, J.; Csikszentmihalyi, M. The Experience sampling method. Sage Publications; Thousand Oaks: 2007.

Hofferth, S.; Flood, S.; Sobek, M. American Time Use Survey Data Extract System: Version 2.4. [Machine-readable database]. Maryland Population Research Center, University of Maryland, College Park, Maryland, and Minnesota Population Center, University of Minnesota; Minneapolis, Minnesota: 2013. http://www.atusdata.org

Huppert FA. Psychological well-being: Evidence regarding its causes and consequences. Applied Psychology: Health and Well-Being. 2009; 1(2):137–164. doi:10.1111/j.1758-0854.2009.01008.x.

Juster, T.; Courant, P.; Duncan, G.; Robinson, J.; Stafford, F. [9 July 2014] Time use in economic and social accounts, 1975–1976. 1975. http://www-2009.timeuse.org/information/studies/data/downloads/usa/1975/ 1975-76-time-use-codebook.pdf.

Kahneman D, Krueger A. Development in the measurement of subjective well-being. Journal of Economic Perspectives. 2006; 20(1):3–24. doi:10.1257/089533006776526030.

Kahneman D, Krueger A, Schkade D, Schwarz N, Stone A. A survey method for characterizing daily life experience—The day reconstruction method. Science. 2004; 306(5702):1776–1780. doi: 10.1126/science.1103572. [PubMed: 15576620]

Krueger A, Kahneman D, Fischler C, Schkade D, Schwarz N, Stone A. Time use and subjective well-being in France and the U.S. Social Indicators Research. 2009a; 93(1):7–18. doi:10.1007/s11205-008-9415-4.

Krueger, A.; Kahneman, D.; Schkade, D.; Schwarz, N.; Stone, A. National time accounting— The currency of life.. In: Krueger, A., editor. Measuring the subjective well-being of nations. University of Chicago; Chicago: 2009b. p. 9-86.

Krueger A, Schkade D. The reliability of subjective well-being measures. Journal of Public Economics. 2008; 92(8–9):1833–1845. doi:10.1016/j.jpubeco.2007.12.015. [PubMed: 19649136]

Kubey, R.; Csikszentmihalyi, M. Television and the quality of life: How viewing shapes everyday experience. Routledge; Hillsdale, NJ: 1990.

Lyubomirsky S, Lepper HS. A measure of subjective happiness: Preliminary reliability and construct validation. Social Indicators Research. 1999; 46(2):137–155.

Marczyk, G.; DeMatteo, D.; Festinger, D. Essentials of research design and methodology. Wiley; Hoboken: 2005.

Michelson, William. From sun to sun: Daily obligations and community structure in the lives of employed women and their families. Rowman and Allenheld; Totowa, NJ: 1985.

Michelson, William. Time use: Expanding explanation in the social sciences. Paradigm Publishers; Boulder, CO: 2005.

National Research Council. Panel on Measuring Subjective Well-Being in a Policy-Relevant Framework. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The National Academic Press; Washington, D.C.: 2012. The subjective well-being module of the American Time Use Survey: Assessment for its continuation..

Nunnally, J. Psychometric theory. McGraw-Hill; New York: 1978.

O'Rourke, N.; Hatcher, L.; Stepanski, EJ. A step-by-step approach to using SAS for univariate and multiple statistics. SAS Institute and Wiley; Cary, NC: 2005.

OECD. Guidelines on measuring subjective well-being. OECD; Paris, France: 2013. http://www.oecdilibrary.org/docserver/download/3013031e.pdf?expires=1393902234&id=id&accname=guest&checksum=D7E49C6C82CC6BE78C9D2DEB29F8B075. [3 March 2014]

Osgood, CE.; Suci, GJ.; Tannenbaum, PH. The measurement of meaning. University of Illinois Press; Champaign: 1957.

Patulny R, Fisher K. Advancing wellbeing research: Would Americans be happier if they lived like Australians? Australian Journal of Social Issues. 2012; 47(1):29–50.

Robinson, J. As we (still) like it: Socializing, religion, kids remain our favourite daily activities. Maryland Population Research Center; 2013. Working paper No. PWP-MPRC-2013-024 http://papers.ccpr. ucla.edu/papers/PWP-MPRC-2013-024/PWP-MPRC-2013-024.pdf. [31 Jan 2014]

Robinson J, Martin S. Comments on Krueger presentation and article. Social Indicators Research. 2009; 93(1):27–30. doi:10.1007/s11205-008-9366-9.

United Nations Economic Commission for Europe. Guidelines for harmonizing time-use surveys. UNECE; Geneva: 2013. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2013/october/17Add1-TimeUseSurvey_Guidelines_UNECE.pdf. [3 March 2014]

Veenhoven, R. World Database of Happiness: Continuous register of scientific research on subjective appreciation of life. Erasmus University Rotterdam; 2006. http://worlddatabaseofhappiness.eur.nl. [11 July 2014]

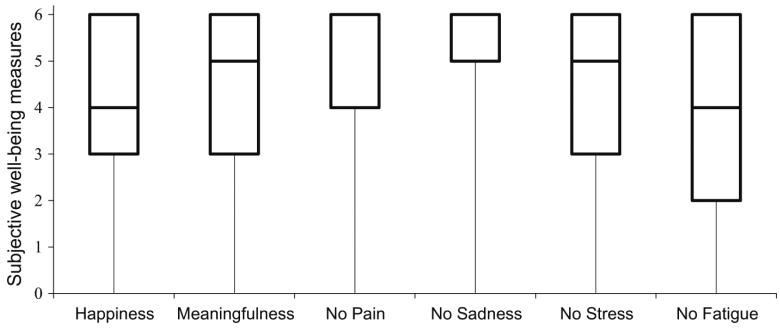

Fig. 1.

Box plot of the subjective well-being measures in the activity-level data. *Source*: The 2010 well-being module of the American Time Use Survey data, own illustration

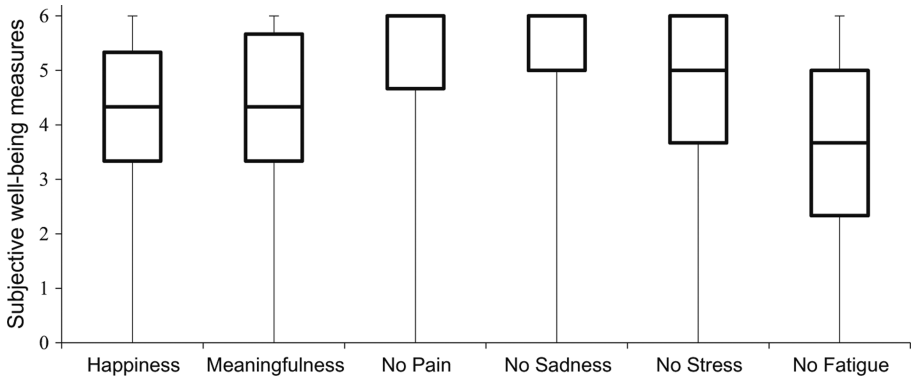

Fig. 2.

Box plot of the subjective well-being measures in the respondent-level data. *Source*: The 2010 well-being module of the American Time Use Survey data, own illustration

Table 1

Weighted means and standard deviations of the subjective well-being measures

	Happiness	Meaningfulness	No pain	No sadness	No stress	No fatigue
Activity-level						
Mean	4.20	4.36	5.02	5.29	4.34	3.65
SD	1.65	1.91	1.66	1.42	1.80	1.98
N	26,091	26,062	26,123	26,123	26,123	26,119
Respondent-level						
Mean	4.23[a]	4.27[a]	5.08	5.34	4.52	3.69
SD	1.35	1.51	1.52	1.22	1.51	1.67
N	8791	8791	8791	8791	8791	8791

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

[a]Not statistically different at the 0.05 level. Each measure ranges from 0 (Not at all) to 6 (Very)

Table 2

Correlation matrix of the subjective well-being measures

	Happiness	Meaningfulness	No pain	No sadness	No stress	No fatigue
Happiness		0.48***	0.21***	0.37***	0.36***	0.22***
Meaningfulness	0.40***		–0.01	0.07***	0.08***	0.03**
No pain	0.17***	–0.00		0.45***	0.40***	0.40***
No sadness	0.32***	0.06***	0.39***		0.60***	0.35***
No stress	0.33***	0.06***	0.34***	0.50***		0.48***
No fatigue	0.17***	0.03***	0.35***	0.27***	0.37***	

Activity-level below the diagonal: Cell sizes vary from 26,023 to 26,123; Respondent-level above the diagonal: Cell sizes are 8791

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

** $p < .01$

*** $p < .001$

Table 3

Highest factor loadings obtained from the subjective well-being measures

One-factor solution	Activity-level		Respondent-level	
	Factor 1 (6 items)	Factor 1 (5 items)	Factor 1 (6 items)	Factor 1 (5 items)
Happiness	0.60	0.55	0.62	0.56
Meaningfulness	0.26		0.27	
No Pain	0.64	0.67	0.67	0.69
No Sadness	0.75	0.76	0.78	0.79
No Stress	0.77	0.78	0.81	0.82
No Fatigue	0.61	0.63	0.67	0.68
Eigenvalue	2.35	2.31	2.61	2.56
% of variance explained	39.21	46.30	43.42	51.26
N	25,989		8791	

Two-factor solution	Factor 1	Factor 2	Factor 1	Factor 2
Happiness		0.78		0.80
Meaningfulness		0.86		0.89
No Pain	0.73		0.74	
No Sadness	0.73		0.77	
No Stress	0.75		0.80	
No Fatigue	0.69		0.72	
Eigenvalue	2.35	1.25	2.61	1.31
% of variance explained	39.21	20.87	43.42	21.79
N	25,989		8791	

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

Table 4

Ordinal logistic regression of general health on the subjective well-being measures and socio-demographic variables

	Model 1 OR (95 % CI)	Model 2 OR (95 % CI)	Model 3 OR (95 % CI)
Subjective well-being measures			
Two factors			
Factor 1	1.81 (1.70, 1.93)***		
Factor 2	1.27 (1.20, 1.34)***		
Individual measures			
Happiness		1.18 (1.12, 1.24)***	
Meaningfulness		1.06 (1.01, 1.10)**	
No pain		1.42 (1.36, 1.49)***	
No sadness		0.95 (0.90, 1.01)	
No stress		1.06 (1.01, 1.11)*	
No fatigue		1.12 (1.07, 1.16)***	
One factor			
Factor 1			1.89 (1.78, 2.01)***
Socio-demographic variables			
Age	0.98 (0.98, 0.99)***	0.98 (0.98, 0.99)***	0.98 (0.98, 0.99)***
Education in years	1.12 (1.10, 1.14)***	1.12 (1.09, 1.14)***	1.12 (1.10, 1.14)***
Family income (ref.: lowest quintile)			
Second quintile	1.05 (0.87, 1.27)	1.05 (0.87, 1.27)	1.05 (0.87, 1.27)
Third quintile	1.20 (0.98, 1.48)	1.23 (1.00, 1.51)*	1.21 (0.98, 1.48)
Fourth quintile	1.40 (1.14, 1.71)**	1.42 (1.16, 1.73)***	1.40 (1.14, 1.71)**
Highest quintile	1.66 (1.32, 2.08)***	1.64 (1.31, 2.06)***	1.67 (1.33, 2.08)***
Employment status (ref.: nonworking)			
Working	1.54 (1.35, 1.76)***	1.51 (1.33, 1.72)***	1.54 (1.35, 1.76)***
Race/ethnicity (ref.: White)			
Black	0.73 (0.62, 0.86)***	0.72 (0.61, 0.85)***	0.73 (0.62, 0.86)***
Asian	0.73 (0.56, 0.96)*	0.69 (0.52, 0.90)**	0.73 (0.56, 0.96)*
Hispanic	0.66 (0.56, 0.77)***	0.61 (0.52, 0.72)***	0.65 (0.55, 0.77)***
Other	0.82 (0.51, 1.31)	0.80 (0.50, 1.30)	0.82 (0.51, 1.31)
Gender (ref.: female)			
Male	0.87 (0.78, 0.97)**	0.88 (0.79, 0.98)*	0.87 (0.78, 0.97)*
Children < age 18 in the household (ref.: No)			
Yes	0.93 (0.82, 1.05)	0.92 (0.81, 1.04)	0.93 (0.82, 1.05)
Marital status (ref.: Unmarried)			

	Model 1 OR (95 % CI)	Model 2 OR (95 % CI)	Model 3 OR (95 % CI)
Married	1.13 (0.99, 1.29)	1.15 (1.01, 1.31)*	1.13 (0.99, 1.29)
Sleep time (divided by 10 min)	1.00 (1.00, 1.01)	1.01 (1.00, 1.01)	1.00 (1.00, 1.01)
Any physical/cognitive difficulty (ref.: yes)			
No	3.41 (2.67, 4.34)***	2.78 (2.18, 3.55)***	3.42 (2.67, 4.36)***
T_1	–4.03***	–7.03***	–4.04***
T_2	–2.26***	–5.22***	–2.26***
T_3	–0.53	–3.44***	–0.53*
T_4	1.47***	–1.39***	–1.46***
Wald Chi square	1201.16***	1432.08***	1201.73***
df	18	22	17
N	8,791	8,791	8,791

This model predicts the probability of being in a lower category (better health) rather than a higher category (worse health)

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

CI: confidence interval; Factor 1 in Model 1: Factor scores obtained from no pain, no sadness, no stress, and no fatigue; Factor 2 in Model 1: Factor scores obtained from happiness and meaningfulness; Factor 1 in Model 3: Factor scores obtained from happiness, meaningfulness, no pain, no sadness, no stress, and no fatigue

* $p < .05$

** $p < .01$

*** $p < .001$

Table 5

Summary of the reliability and validity tests

	Option 1	Option 2	Option 3	Option 4
Number of items	2	4	5	6
Reverse-coded negative feelings				
No pain		X	X	X
No sadness		X	X	X
No stress		X	X	X
No fatigue		X	X	X
Positive feelings				
Happiness	X		X	X
Meaningfulness	X			X
Internal consistency				
Average inter-item correlation	$0.40_{(a)}$-$0.48_{(r)}$	$0.37_{(a)}$-$0.45_{(r)}$	$0.32_{(a)}$-$0.43_{(r)}$	$0.25_{(a)}$-$0.30_{(r)}$
Average factor loading	$0.82_{(a)}$-$0.85_{(r)}$	$0.73_{(a)}$-$0.76_{(r)}$	$0.68_{(a)}$-$0.71_{(r)}$	$0.61_{(a)}$-$0.64_{(r)}$
% of variance explained	$20.9_{(a)}$-$21.8_{(r)}$	$39.2_{(a)}$-$43.4_{(r)}$	$46.3_{(a)}$-$51.3_{(r)}$	$39.2_{(a)}$-$43.4_{(r)}$
Cronbach's alpha	$0.59_{(a)}$-$0.65_{(r)}$	$0.68_{(a)}$-$0.74_{(r)}$	$0.69_{(a)}$-$0.74_{(r)}$	$0.66_{(a)}$-$0.70_{(r)}$
Validity				
Correlation with health	0.07	0.32	0.33	0.33
Odds ratio	1.27	1.81	1.89	1.89

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

(a) Indicates internal consistency estimates obtained in the activity-level data and (r) indicates those obtained in the respondent-level data. All validity estimates are calculated in the respondent-level data

Table 6

Coefficients of variation in the subjective well-being measures

	Happiness	Meaningfulness	No pain	No sadness	No stress	No fatigue
Activity-level						
CV (%)	39.29	43.81	33.07	26.84	41.47	54.25
N	26,091	26,062	26,123	26,123	26,123	26,119
Respondent-level						
CV (%)	31.91	35.36	29.92	22.85	33.41	45.26
N	8791	8791	8791	8791	8791	8791

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

Table 7

Correlations of the subjective well-being measure across activities

	First activity	Second activity	Third activity
First activity			
Second activity	0.70***		
Third activity	0.63***	0.69***	
Mean	4.51	4.45	4.50
SD	1.05	1.07	1.06
CV (%)	23.28	24.04	23.56
N	8791	8772	8573

Each of the mean measures ranges from 0 (not at all) to 6 (very)

Source: The 2010 well-being module of the American Time Use Survey data, own calculation

*** $p < .001$