

# The Cornerstone of Research on Subjective Well-Being: Valid Assessment Methodology

By William Pavot, Southwest Minnesota State University

## Citation:

Pavot, W. (2018). The cornerstone of research on subjective well-being: Valid assessment methodology. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being*. Salt Lake City, UT: DEF Publishers. DOI:nobascholar.com

## Abstract:

This chapter focuses on the process of obtaining valid assessments of Subjective Well-Being (SWB) from individuals, primarily with conventional self-report methodology. A brief review of the SWB construct and its constituent facets is followed by a discussion of some of the factors, both stable and transient, that may influence an individual's response to questions about their SWB. Several examples of measurement instruments, ranging from single-item measures to multi-faceted questionnaires, are briefly reviewed. The issues of scale reliability and validity are considered. General design strategies for optimizing the "fit" between assessment and the goals of the research follow. Some conclusions regarding SWB research, as well as additional resources and guides that are available to researchers new to SWB assessment are in a final section.

**Keywords:** Subjective Well-Being, Assessment, Self-report, Reliability, Validity

---

Research on Subjective Well-Being (SWB) continues to be a thriving inter-disciplinary interest area. Even a brief examination of the chapter topics in this volume confirms the breadth and depth of SWB-related investigation. This level of interest is quite remarkable for a topic that was almost unheard of three decades ago. In the interim, an impressive database has been gathered in an attempt to discern the structure, correlates, and sources of SWB.

Theoretical research focused on the structure and sources of SWB continues, but the application of SWB research for the improvement of the quality of life has become an important additional growth area. Many nations, for example, have begun to recognize that the assessment of the SWB of their citizenry can provide an important social indicator for policy and decision making, over and above the more traditional sources, such as economic indicators (Diener, 2000; 2006).

Despite the rich diversity of research related to SWB, some concerns are common: How accurate are the observations upon which the conclusions of the investigator are based? What are the potential strengths and weaknesses of the research design employed? Are the findings widely generalizable? Regardless of the goals of the researcher, the fundamental building block continues to be dependable and accurate data obtained from valid assessment methodologies. The present chapter will examine some of the issues involved in the assessment of SWB, with a specific focus on traditional self-report measures.

The first section of this chapter will briefly discuss the major components of the SWB as they have emerged from early research. The second section will review and discuss some of the measurement issues that may influence the validity of self-reports of SWB. Examples of some of the best-known instruments for the assessment of SWB are presented and discussed in the third section of the paper, and, in the fourth section, issues of establishing the reliability and validity of assessment instruments are considered. Some design considerations for future assessments of SWB appear in the following section, and some conclusions and suggested resources are presented in the closing section.

## The Structure of SWB

The structure of SWB can be conceptualized as having two, three, or four facets, depending on the

specificity required to address the research question(s) of interest. At a very fundamental level, SWB can be divided into two facets: One facet would include the affective or emotional aspect of subjective experience, and a second facet that would represent the cognitive, evaluative or reflective aspect. This structure is intuitively appealing and straightforward. But early findings by Bradburn (1969), and later confirmed by Diener and Emmons (1984) indicated that the experience of positive emotion and negative emotion are largely independent of each other. As a result, most contemporary researchers tend to consider positive emotion/affect and negative emotion/affect as separate facets, and tend to use a three-facet model of SWB (e.g., Arthaud-Day, Rode, Mooney, & Near, 2005). Additionally, depending on the goals of the researcher, the cognitive/evaluative/reflective component may sometimes be further divided as well. Frequently, investigators are interested in an individual's overall evaluation of their life as a whole, usually referred to as life satisfaction. However, if the research is focused on a particular aspect of a person's life, such as their employment or their marriage, it may be useful to examine the individual's evaluation of a particular domain of their life. Thus, the evaluative component can be divided into a judgment of overall life satisfaction, or one or more specific aspects of life with domain satisfaction.

### **Issues in the Measurement of Subjective Well-Being**

On the surface, the assessment of SWB may seem (perhaps deceptively) simple. In many cases, a single, clear question is asked, such as "Overall, how happy are you in life?" The respondent is typically asked to make a rating of their happiness using a Likert-type response scale (e.g., ranging from "1", indicating a low level of happiness to a "10" indicating a very high level of happiness). The researcher's assumption is that the response provided is an accurate reflection of the respondents overall and relatively stable subjective experience. To an extent, a very large body of empirical evidence has supported the validity of this assumption. But previous research has also revealed that a number of factors can, under specific circumstances, exert influences on the response given, possibly biasing the response in a positive or negative direction, and thereby reducing the validity of the data obtained.

**Random transient influences on self-reports of SWB.** A potential concern to researchers attempting to assess SWB is the influence that relatively random, contextual and/or situational factors might have on an individual's response. These transient influences might include current mood, the situation surrounding the assessment situation, or the influence of questions or items that are presented prior to the SWB item(s) within a given questionnaire arrangement.

**Transient affective/mood states.** One well-known demonstration of the potential for current mood to influence self-reports of SWB was provided by Schwarz and Clore (1983). Using induced positive or negative mood states (e.g., memory searches for good or bad events; obtaining responses on sunny versus rainy days) these researchers were able to influence respondents judgements of overall happiness and satisfaction with life. Subsequent research (Pavot & Diener, 1993a; Eid & Diener, 2004; Lucas & Lawless, 2013; Yap et al., in press) has shown that these effects are generally minimal, if not nil, and can be further reduced with methodological safeguards; nonetheless, the potential influence of current mood on self-reports of SWB, while generally small, is noteworthy.

**Transient situational/contextual factors.** It is also possible for self-reports of SWB to be influenced by transient factors in the situation or the context surrounding the response. A face-to-face interview, for example, could yield reports of greater SWB, when compared to responses on an anonymous survey, due to the effects of the social desirability of happiness. In situations where an item intended to measure SWB is embedded in a larger survey, the content areas of the preceding items in the survey might influence the response to the specific SWB question. The order in which questions are presented can have an influence on an individual's response (Strack, Martin, & Schwarz, 1988).

While it is important to recognize and acknowledge that these random transient factors can influence self-reports of SWB; it is also important to point out that these influences are generally limited, and can be minimized with careful design and methodology. In very large survey designs, for example, transient mood effects are likely to have minimal influence on group levels of reported SWB. Barring some major crisis or world event with sweeping emotional impact, the effects of the transient negative mood state of one individual is likely to be effectively cancelled by the transient positive mood state of another. The effects of item placement on response can be minimized by presenting survey items in a random order. In smaller scale studies specifically focused on an aspect of well-being, transient effects can be minimized by assessing SWB on more than one occasion, and then computing an average score across those occasions. While some transient effects will likely always be part of self-reported SWB, careful methodological planning can substantially reduce their impact.

**Stable situational /contextual factors.** In addition to the potential influences of transient situational and contextual factors may exert on self-reports of SWB, investigators must also be mindful of more stable influences as well. A differential cultural perspective between respondents, for example, might

well be a source of consistent variation in assessments of SWB. Although several studies (e.g., Balatsky & Diener, 1993; Scollon, Diener, Oishi, & Biswas-Diener, 2004) have indicated that SWB measures do have a reasonable level cross-cultural validity, other reports (e.g., Vitterso, Roysamb, & Diener, 2002) have indicated that the reliability and underlying factor structure of life satisfaction measures can vary across cultures. Examining nations using a general individualist versus collectivist cultural categorization can provide an illustration. Using two large international data sets, Suh, Diener, Oishi and Triandis (1998) found life satisfaction more strongly correlated with emotions in individualist nations as compared to collectivist nations. In collectivist nations, social norms were equally as strong as emotions in the prediction of life satisfaction reports. Thus the sources of information that are accessed in the formulation of life satisfaction judgments can vary across cultures.

An additional factor, less clearly identified yet likely no less influential, is the potential for differential “cohort effects” between age groups (Diener & Suh, 1997). Individuals of the same chronological age (same “cohort”) share not only age-related physical developmental changes as they progress through life; they have also experienced social change and major historical events together. These shared experiences likely serve to help each cohort group develop its own “scale” of positive and negative life events that is to some extent unique. For example, a cohort that has experienced severe economic depression or a world-wide conflict likely has a somewhat different scale of negative events than members of a cohort without these experiences.

Another factor that can influence responses to SWB measures is the particular facet of SWB that is assessed by the measure. Measures of the evaluative facet of SWB (i.e., judgments of life or domain satisfaction) appear to be influenced to a greater degree by contextual factors than measures focused on the affective facets of SWB (Tay, Chan, & Diener, 2014).

### **Memory Issues and Biases**

Another potential influence on self-reports of SWB involves the memory processes used by the respondent in formulating their subjective report. “Traditional” self-report instruments for assessing SWB rely on global retrospective reports from respondents. Unfortunately, human memory is not perfect, and a number of factors may work to create biases in retrospective reports. For example, the data indicate that people tend to formulate reports of emotion that are more strongly correlated with the amount of time they have experienced an emotion, rather than on the intensity of the emotional experience (Diener, Sandvik, & Pavot, 1991).

One methodological innovation intended to address the problems inherent in retrospective reports is the Experiential Sampling Method (ESM; Csikszentmihalyi & Larson, 2014). This method involves obtaining a large number of reports of emotion and SWB at random moments during each day, while the respondents are having these subjective experiences. Such an approach can eliminate the retrospective element of the self-report, and therefore presumably eliminate the problem of memory bias. The ESM method is discussed in detail in another chapter of the volume.

### **Representative Examples of SWB Self-Report Instruments**

Rutt Veenhoven has created a large and highly informative database of SWB research, the World Database of Happiness. One section of this database focuses on existing measures of happiness, which number more than eleven hundred catalogued examples (Veenhoven, 2017). It is not possible to review these measures in detail; it is possible, however, to sort potential instruments into general groups, and prominent examples of each general group are reviewed in this section.

**Single-item broadband measures.** Most of the measures that have been included and catalogued in the World Database of Happiness are single-item measures, usually presented to respondents in a survey format (Veenhoven, 2017). Many of these single-item measures are “broadband” in design; all facets of SWB are assessed by this single item, in the form of a statement or question. Usually the stimulus item is focused on the generic terms of “Happy” or “Happiness.” A prototypical example of this category is the Fordyce’s Single-Item Happiness Question (1977). The Fordyce Scale offers 11 Likert-scale response choices (ranging from “Extremely happy” to “Extremely unhappy”) to the query: “In general, how happy or unhappy do you usually feel?”

Another single-item broadband measure of SWB is the Delighted-Terrible Scale (Andrews & Withey, 1976). After prompting the respondent to think about their own life and about “life in this country” over the past year, and their expectancies for the future, the respondent is asked to make one of seven possible responses to the question: “How do you feel about how happy you are?” The responses range from 1 (Terrible) to 7 (Delighted). The Delighted-Terrible scale has been used in conjunction with the World Values Survey (1994), and has demonstrated good psychometric characteristics.

The brevity of single-item broadband measures is obviously a great asset to survey researchers.

But there is a significant trade-off involved; single item measures do not provide specific information regarding the separate facets of SWB.

**Single-item facet-focused measures.** One of the most enduring and widely used single-item measures is Cantril's Self-Anchoring Ladder (SAL; Cantril, 1965). This instrument has been frequently used in large-scale survey studies, perhaps most notably research conducted by the Gallup Organization, both internationally (Helliwell, Layard & Sachs, 2014) and in the United States (Harter & Gurley, 2008). Data from the SAL is a primary source for the World Happiness Report, sponsored by the Organisation (sic) for Economic Co-operation and Development (OECD). The instructions for the Cantril Self-Anchoring Scale ask the respondent to imagine a ladder with the steps numbered from zero (the bottom step) to 10 (the top step). Then the top and bottom step are identified as representing the best possible life and worst possible life, respectively, for the respondent. The respondent then chooses the step that is most representative of their life as a whole. The time frame for the response can be adjusted to represent their current life, their future life (i.e., five years from now), or both time-frames can be assessed separately. Single item measures of Life Satisfaction have been demonstrated to have good reliabilities (Cheung & Lucas, 2014), that are comparable to multiple-item scales.

**Multi-item comprehensive scales.** In some contrast to the single-item survey measures of SWB, which are often used for a broad range of purposes and applications, multiple-item measures tend to be used in research settings with a primary focus on SWB. Multiple-item measures can provide researchers with a more detailed picture of the respondent's SWB, and can be examined in more detail psychometrically.

Most multi-item SWB scales are facet specific; they tend to assess either the affective facets *or* the life satisfaction facet of SWB, but not both. Still, there are examples of multi-item, comprehensive measures of SWB.

The Oxford Happiness Questionnaire (OHQ; Hills & Argyle, 2002), represents a revision of an earlier instrument, the Oxford Happiness Inventory (OHI; Argyle, Martin, & Crossland, 1989). The OHQ consists of 29 items, each a statement to which individuals can respond on a 6-point Likert-type scale. The authors interpret the underlying factor as being uni-dimensional. Correlational comparisons indicate that both the convergent and discriminant validity of the OHQ are somewhat improved over the OHI.

The Subjective Happiness Scale (SHS; Lyubomirsky & Lepper, 1999) is a brief (4-item) measure that is also intended to assess global subjective happiness. Responses to the four items can be made on a 7-point Likert scale, with higher numbers indicating greater SWB. In early validation studies, correlations of the SHS with informant reports were moderate to strong, and test-retest reliability was also strong. The brevity of the SHS, combined with the encouraging initial psychometric data on the scale, suggest that the scale should have good utility when a comprehensive measure is desired.

**Multi-item affect scales.** Several multiple-item measures have been developed with a specific focus on the affective components of SWB. Although it would be possible to create a scale that was focused on either positive affect or on negative affect exclusively, most of the scales in this category measure both of these facets using separate subscales. This allows for the computation of an index of the affective state of the respondent, commonly referred to as "affect balance" (Bradburn, 1969). The affect balance score represents the difference between the amount of positive affect reported and the amount of negative affect reported (e.g., positive affect – negative affect = affect balance). A positive affect balance would indicate that the respondent has reported a greater amount of positive emotion than negative emotion over a specified time frame and presumably has experienced SWB during that time. It is also possible, of course, to focus on the subscale scores independently as well.

The Bradburn Affect Balance Scale (BABS; Bradburn & Caplovitz, 1965; Bradburn 1969) was the first scale to use the affect balance approach as an assessment of SWB. The 10-item BABS includes PA and NA subscales of five items each. In the original version of the BABS, the respondent is asked to indicate, using a "yes" or "no" response, whether they have experienced any of the positive or negative affects "during the past few weeks." The BABS affect balance index is generally correlated (albeit moderately) to other SWB indices.

The Affectometer 2 (Kammann & Flett, 1983) uses an approach that is similar to the BABS, but incorporates an expanded array of 40 affect items, intended to assess ten facets of SWB. The Affectometer 2 also replaces the "yes" or "no" responses of the BABS with a frequency response scale for each item. The combination of larger number of items and a frequency response scale are psychometrically appealing, but the length of the Affectometer 2 might be an obstacle in some research designs.

Researchers have frequently used another multiple-item affect measure, the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), in recent years. The PANAS includes 10

affective adjectives in each of its two subscales. Respondents indicate the degree to which they have experienced each emotion, using a 5-point scale. The time-frame can be adjusted to reflect either the current emotional experience of the respondent or a retrospective report (e.g., “over the past few days” or “in the past few weeks”). Both the PA and NA subscales have good psychometric characteristics. The PANAS tends to put emphasis on high activation emotions, and includes some adjectives (e.g., “strong”) that may not necessarily reflect emotion exclusively. An expanded version of the original PANAS, the PANAS-X, utilizing 60 items, is also available (Watson & Clark, 1999).

A relatively recently developed scale, the Scale of Positive and Negative Experience (SPANE; Diener et al., 2010), has a number of desirable characteristics. The 12 items (emotion adjectives) of the SPANE, despite the brief format, assess a broad range of both positive and negative affective experience. The 5-point frequency response scale, ranging from 1 = “Very rarely or never” to 5 = “Very often or always” indicates the relative amount of time that the respondent was experiencing each emotional state. And frequency of feelings, rather than intensity of emotion, tends to be related to other SWB measures, such as life satisfaction measures.

**Multi-item satisfaction scales.** Probably the best known of the measures in this category is the Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985). The satisfaction with life scale has been used in thousands of studies and has been translated into more than 30 languages. The SWLS has been cited in more than 17 thousand published scholarly articles (according to a 2017 Google Scholar search). The SWLS consists of five statements, such as “In most ways my life is close to my ideal,” and “I am satisfied with my life.” Typically, a seven-point Likert scale is provided for a response to each item, ranging from “strongly disagree” (1) to “strongly agree” (7). The five-item SWLS has demonstrated good internal reliability and moderate temporal reliability. It has been shown to be sensitive to both positive and negative change over time (Pavot & Diener, 1993b; Pavot & Diener, 2008).

The 15-item Temporal Satisfaction With Life Scale (TSWLS; Pavot, Diener, & Suh, 1998) uses essentially the same items as the SWLS, but presents them with three distinct temporal contexts, referring to the past, the present, and the expected future satisfaction of the respondent. The TSWLS thus can be used to gauge future expectancies of the respondent, as well as past and current perspectives.

**Population / domain specific measures.** Nearly all of the measures of SWB reviewed in this chapter have good generalizability to a very wide range of populations. Most will provide valid SWB data without concern for respondent age, or occupational, economic or other noteworthy social differences. However, some researchers are interested in aspects of life that may be particularly relevant to the SWB of a specific group or age range. Specialized measures of SWB and life satisfaction are available for a number of specific populations. For example, researchers interested in the aging process have developed several measures, such as multiple versions of a Life Satisfaction Scale (LSR, LSIA, LSIB; Neugarten, Havighurst, & Tobin, 1961). Another well-known scale intended for older adults is the Philadelphia Geriatric Center’s Morale Scale (PGC Morale Scale; Lawton, 1972, 1975). This scale appears to assess several factors, including loneliness and attitudes toward aging. These scales are well-established and have been used extensively in research focused on late adulthood.

Children and adolescents represent another population that has been the focus of attention for well-being researchers, particularly in their role as students. The Student's Life Satisfaction Scale (Huebner, 1991) and the Brief Multidimensional Students' Life Satisfaction Scale (Seligson, Huebner, & Valois, 2003) are two examples of measures focused on the student population.

Many researchers are interested in assessing satisfaction with specific aspects of life, such as job, marriage, income, religious orientation, government function, or some other specific life domain. Items intended to assess specific domains are numerous (Veenhoven, 2017), and can be adjusted to whatever situation or domain is of interest to the researcher.

**Measures of constructs related to SWB.** When designing a study focused on SWB, it is often desirable to include additional assessment instruments that provide data on constructs that are related or are believed to be related to SWB itself. An example of one such construct that has relevance for the experience of SWB is Dispositional Optimism. Dispositional Optimism refers to a global generalized tendency to believe that one will generally experience good vs bad outcomes in life (Scheier & Carver, 1985). One measure intended to assess dispositional optimism is the Life Orientation Test (L.O.T.; Scheier & Carver, 1985). The L.O.T. includes eight construct-relevant items (four keyed in a positive direction, and four in a negative direction), along with four “filler” items intended to disguise the specific focus of the scale. Responses are made on a five-point Likert scale. A later version of the scale, the 10-item L.O.T. – R., was developed by dropping two of the construct-relevant items, in order to more clearly distinguish dispositional optimism from neuroticism (Scheier, Carver, & Bridges, 1994). Both versions have good psychometric characteristics, and have been used extensively in research.

The Flourishing Scale (FS; Diener et al., 2010) is intended as a broadband measure of self-perceived well-being across a range of important domains. The FS assess functioning in areas that are key to psychological well-being such as relationships, feelings of competence, and a sense of meaning and purpose in life. The FS includes eight items, and utilizes a 7-point Likert scale for responses.

Another recently developed broadband measure also focused on a range of well-being related domains is the Comprehensive Inventory of Thriving (CIT; Su, Tay, & Diener, 2014). The CIT is composed of 54 total items (18 subscales), and spans a very broad range of domains. For situations requiring a quicker assessment, a 10-item version of the CIT, The Brief Inventory of Thriving (BIT), has also been developed and is included in the same report. Both of these measures were evaluated with multiple samples with diverse demographics. The CIT and BIT demonstrated excellent psychometric properties and good convergent validity (Su, Tay, & Diener, 2014)

### **Reliability and Validity of SWB Assessments**

**Reliability.** Reliability refers to the degree to which an assessment instrument produces consistent, dependable results. Several forms of reliability are commonly examined when evaluating an assessment instrument or technique. For the present I will focus on three basic types of reliability. One gauge of reliability is focused on the ability of assessment instrument to produce dependable results over a period of time, and is often referred to as “test-retest” reliability. For example, if we asked an individual to complete a measure of SWB today, and then asked them to complete the same measure one month from now, we would expect that the results would be similar. The degree of agreement or correlation between the two measurement occasions would represent an index of test-retest reliability. A second form of reliability is referred to as “inter-rater” reliability. This form of reliability is an index of agreement between two or more “raters” or “judges.” For example, we might ask two (or more) different raters to evaluate the emotion content of a writing sample that a respondent has completed. The degree to which their evaluations agree would serve as an index of inter-rater reliability. A third type of reliability is usually identified as internal reliability or “Alpha” reliability. The “Alpha” identifier is a reference to Cronbach’s Alpha (Cronbach, 1951), which is a statistic that is usually computed and reported as an index of internal reliability. Internal reliability is determined by the consistency of response to each of the items on a multiple-item measure. That is, the tendency of individuals to respond consistently to all the items of the measure in a similar way. Thus, the greater the consistency of response across the individual items of the scale, the larger the Alpha statistic will be. There is no absolute standard regarding the minimum acceptable level of Alpha; generally a range of .7 to .75 would be considered a minimal demonstration of internal consistency.

The reliability of most self-report measures of SWB is typically evaluated using either an index of test-retest reliability or of internal reliability. For multiple-item measures, it is usually expected that statistics depicting both indices will be presented. With single-item measures, the index of internal reliability, or Cronbach’s Alpha, cannot be computed. In either case, an index of reliability is a critical element in the evaluation of the assessment instrument or method.

**Validity.** Validity can also be understood from multiple perspectives. Many types of validity have been identified, such as face validity and predictive validity. A full discussion of all forms of validity is beyond the scope of this chapter; it is perhaps more efficient to focus on some central elements. A central form of validity, construct validity, refers to the degree to which an instrument accurately measures the construct it is intended to measure. In other words, a measure of SWB, when completed by a respondent, should be an accurate depiction of that person’s experience. Establishing the construct validity of a SWB measure is a complex process. A first step in the process usually involves establishing reliability, because without dependable assessment, it is impossible for a measure to be considered valid. Thus, reliability is a necessary, but not sufficient, condition for validity. Reliability is not sufficient for validity because, unfortunately, it is possible to be reliably wrong. In addition to demonstrating reliability, it is critical that an assessment instrument demonstrate that it is accurately measuring the construct it is intended to measure. Accuracy (construct validity) can be evaluated by comparing the results yielded by the measure to the results of both similar and dissimilar measures obtained from the same sample.

**Convergence of similar measures.** In order to go beyond reliability and establish the accuracy (validity) of our measurement, we have to rely on established “benchmarks” that can help us determine how well our measure is accurately assessing the intended construct. One type of benchmark is provided by previously established indices that measure the same construct, or at least similar constructs to the one we are interested in. We would expect that such indices would be positively correlated with our measure. To demonstrate such correlations would represent establishing convergent validity. To gauge the convergent validity of a SWB measure, it is necessary to include more than one such measure within a particular sample, and then examine the strength of the correlations between the two (or more) measures of SWB. Measures of other closely related constructs can also be examined. Thus, a measure of flourishing

would be expected to be positively correlated to a measure of life satisfaction from the same sample.

An additional advantage of including more than one measure of SWB within a research design is the possibility of creating “composite” SWB scores. That is, if the convergence between separate measures of SWB is high, it is possible, using standard scores such as “Z” scores, to combine the results of the individual measures into one composite measure of SWB. The advantage to this strategy is that a composite score or index tends to have greater stability and reliability than the scores on the individual scales, thus strengthening the results of any further analysis. A potential drawback to this strategy, however, is that the use of composite scores can obscure meaningful differences between measures at the specific facet level. It is desirable, therefore, to report the results of basic analyses of the individual scales before combining them into composite scores.

***Convergence of self-reports with other methodologies.*** In addition to demonstrating the convergence of multiple self-reports of SWB, convergence can also be evaluated by comparing self-reports of SWB to an index of SWB obtained by using another methodology. An often-used alternative methodology is the use of “informant” reports. Informant reports are typically obtained from someone who knows the research participant well, such as a spouse, parent, sibling, roommate, or co-worker. Informant reports tend to increase confidence in the validity of the self-reported data, as they are obtained from an independent source. The validity of self-reported measures of SWB, such as reports of life satisfaction, has been demonstrated by the convergence of such reports with informant reports in a number of studies (e.g., Costa & McCrae, 1980; Pavot, Diener, Colvin, & Sandvik, 1991; Pavot & Diener, 1993b; Lyubomirsky & Lepper, 1999).

Other alternative methodologies include rating based on clinical interviews, ratings based on writing samples (Danner, Snowdon, & Friesen, 2001), and analysis of memory bias for positive and negative events (Sandvik, Diener, & Seidlitz, 1993).

***Divergence of dissimilar measures.*** Another set of benchmarks, in this case measures of dissimilar constructs, could also support the validity of our measure. By demonstrating divergent validity, we can show that our measure is an assessment of a construct that is distinct from other, dissimilar constructs. In the case of divergent validity, we would expect to see either no correlation, or perhaps a negative correlation in the case of a strongly dissimilar (perhaps opposite) construct. For example, a measure of SWB would be expected to have a negative correlation with a measure of psychological distress, such as depression or anxiety. Two often used measures of distress that can be useful in this capacity are a depression scale, the CES-D scale (Radloff, 1977), and the SCL-90-R (Derogatis, 1992), a broader assessment of nine primary symptom dimensions. Scores on measures of anxiety and neuroticism, such as the neuroticism scale on the NEO-PI (Costa & McCrae, 1980).

### **Considerations for Designing Future Studies**

For the researcher contemplating a study focused on SWB, the design and methodological possibilities are many. Clearly, a wide array of self-report instruments for the assessment of SWB is available. Choosing a measure, or a battery of measures, and the specific method of employing these measures to best advantage are critical questions. The question of which, if any, additional dimensions to add to a design is important as well. It may be useful to review some of these considerations.

In its most basic form, the assessment of SWB might be accomplished by a single question at one point in time, such as an item embedded in a larger survey of social attitudes. Despite the simplicity of this approach, useful data, albeit at a very general level, can be obtained in this way. This might be the only feasible choice for very large -scale survey efforts, and has been used quite successfully in many such situations, such as compiling the World Happiness Report. The difficulty with this approach is the lack of specificity and detail in the assessment. As these surveys are generally not given to the same individuals at an additional occasion, we have no sense of the stability or the dynamics of the respondent’s subjective experience, nor do we know the specifics of the quality of the subjective experience. Still, the very general snapshot of SWB might well be useful in monitoring the overall SWB of a large group, such as monitoring the overall reaction to a recent change in public policy.

For those interested in a more finely grained assessment, more sophisticated measures may be in order. Two or more SWB measures might be used in concert, and perhaps some additional variables (e.g., personality factors, measures of distress) could be included, to provide a system of benchmarks, to examine convergence or divergence of the measures. If the SWB measures are multiple-item scales, information on internal reliabilities can also be obtained.

Another level of complexity can be added to the analysis if the research participants are surveyed at more than one point in time, with an interval of one or more months between. With assessment on multiple occasions, we can obtain test-retest reliability data, and look for stability (or change) over time. For



example, a meta-analysis of such data has shown test-retest reliabilities for global SWB tend to range from .70 over a one-year period, and then decrease to approximately .50 for a 5-year interval (Anusic, & Schimmack, 2016). If the entire original sample cannot be retested, it may be possible to conduct a retest on a subsample of the original group.

A still more sophisticated design might include the collection of informant reports, in addition to the self-report measures of SWB. Informant reports that show good convergence with the self-reports of the study serve to greatly increase the confidence in the findings of the research. A detailed discussion of the sources of self-report and informant-report agreement is available in a meta-analysis by Schneider and Schimmack (2010).

Thus, a longitudinal research design with multiple measures of SWB assessed on multiple occasions and including informant reports is likely to be desirable for those who want a complex, sophisticated data set of the first order. Obviously the cost to achieve that goal, in terms of time, resources, and effort is going to be high. Conversely, a single-item survey assessment can likely be obtained quickly and economically, but will only provide a very general level of data. For many researchers, a design that falls somewhere between these points is a likely solution. A gradual validation of a finding from a series of studies is as valuable (if not more valuable) than a single “answers all the questions” effort. Replication of previous findings is usually not sensational, but it is essential for building a strong empirical database. A programmatic body of research, in which each study replicates and builds upon previous work, is a valuable tool for advancement.

### **Summary and Additional Resources**

Fortunately, there are many additional sources and examples of the topics briefly presented here. I will list a few of the many options here; there are far too many to enumerate in any comprehensive way.

For those who are new to the study of SWB, and interested in obtaining some general background information, a landmark review article by Ed Diener is a great start (Diener, 1984). After reading the 1984 paper (if you haven't already) I would suggest Diener (2000) as an update, with a particular proposal for a national index of SWB. An edited volume (Kahneman, Diener, & Schwarz, 1999) also provides much greater depth on many of the issues raised in this chapter. For a review focused on the assessment of SWB, Diener (1994) is a good choice; Pavot (2008) also discusses assessment issues. A more recent discussion of the validity of life satisfaction scales is available in Diener, Inglehart and Tay (2013).

For those readers interested in developing and validating their own measures, it may be valuable to review the work of others who have previously engaged in the process. For some good examples of the scale validation process, I would suggest Lyubomirsky and Lepper (1999), Seligson, Huebner, & Valois, (2003) or Diener et al. (2010). The article by Pavot, Diener, Colvin, & Sandvik (1991) is an example of validation of a self-report scale with informant reports. Before you begin the complex process of developing and validating a new measure of SWB, I would encourage you to access the World Database of Happiness (Veenhoven, 2017) on the internet (website: <http://worlddatabaseofhappiness.eur.nl>). This is an outstanding resource for anyone interested in any aspect of research on happiness or SWB. As noted earlier in this chapter, more than eleven hundred measures have been catalogued and are listed in the “measures” section of the database. It is possible that a previous researcher has already developed a measure that may well suit your needs. Guidelines for researchers interested in using SWB assessments for specific purposes (e.g., general assessment of well-being at the national level) are available (Diener, 2006).

There is no single “correct” way to assess SWB across all research situations and purposes. Nevertheless, it is very possible to optimize the validity of assessment in any given study by a careful consideration of the parameters of the research situation, the resources available, and the proximal and distal goals of the particular research program.

It is important that future researchers continue to pursue both theoretical and applied questions regarding SWB. The topics listed in the table of contents of this handbook are indicative of the many possible directions a program of research might take. At the basic research level, questions from the evolutionary, biological, cognitive, and social perspective are the subject of current study, and will continue to be so for the future. Applied research ranges from the development of interventions at the individual level to social policy at the national level. In terms of the assessment of SWB, future research could focus on the further development of alternative methodologies to compliment the already extensive array of self-report measures that are available.

### **References**

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: America's perception of life quality*. New York: Plenum Press.



- Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of Personality and Social Psychology*, *110*(5), 766–781.
- Argyle, M., Martin, M., & Crossland, J. (1989). Happiness as a function of personality and social encounters. In J. P. Forgas, & J. M. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 189-203). North-Holland: Elsevier.
- Arthaud-Day, M. L., Rode, J. C., Mooney, C. H., & Near, J. P. (2005). The subjective well-being construct: A test of its convergent, discriminant, and factorial validity. *Social Indicators Research*, *74*, 445-476.
- Balatsky, G., & Diener, E. (1993). Subjective well-being among Russian students. *Social Indicators Research*, *28*(3), 225-243.
- Bradburn, N. M. (1969). *The structure of psychological well-being*. Chicago: Aldine.
- Bradburn, N. M., & Caplovitz, D. (1965). *Reports of happiness*. Chicago: Aldine.
- Cantril, H. (1965). *The pattern of human concerns*. Rutgers University Press: New Brunswick.
- Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research*, *23*(10), 2809-2818.
- Costa, P. T., Jr., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, *38*, 668-678.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *31*, 93-96.
- Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35-54). Netherlands: Springer.
- Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology*, *80*(5), 804-813.
- Derogatis, L. R. (1992). *SCL-90-R: Administration, scoring and procedures manual for the R (revised) version and other instruments of the Psychopathology Rating Scale Series*. Clinical Psychometric Research.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542-575.
- Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research*, *31*(2), 103-157.
- Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, *55*, 34-43.
- Diener, E. (2006). Guidelines for national indicators of subjective well-being and ill-being. *Applied Research in Quality of Life*, *1*, 151-157.
- Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47*, 1105-1117.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*, 71-75.
- Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research*, *112*(3), 497-527.
- Diener, E., Sandvik, E., & Pavot, W. (1991). Happiness is the frequency, not the intensity, of positive versus negative affect. *Subjective well-being: An interdisciplinary perspective*, *21*, 119-139.
- Diener, E., & Suh, M. (1997). Subjective well-being and age: An international analysis. *Annual Review of Gerontology and Geriatrics*, *17*, 304-324.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*, 143-156.
- Eid, M., & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research*, *65*, 245-277.
- Fordyce, M. W. (1977). *The Happiness Measures: A sixty-second index of emotional well-being and mental health*. Unpublished manuscript. Edison Community College, Ft. Myers, FL.
- Harter, J. K., & Gurley, V. F. (2008). Measuring well-being in the United States. *Association for Psychological Science Observer*, *21*(8).

- Helliwell, J. F., Layard, R., & Sachs, J. (2014). *World happiness report 2013*. New York: UN Sustainable Development Solutions Network.
- Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: A compact scale for the measurement of psychological well-being. *Personality and Individual Differences*, 33(7), 1073-1082.
- Huebner, E. S. (1991). Initial development of the student's life satisfaction scale. *School Psychology International*, 12(3), 231-240.
- Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.
- Kammann, R., & Flett, R. (1983). Affectometer 2: A scale to measure current level of general happiness. *Australian Journal of Psychology*, 35(2), 259-265.
- Lawton, M. P. (1972). The dimensions of morale. In D. Kent, R. Kastenbaum, & S. Sherwood (Eds.), *Research, planning and action for the elderly* (pp. 144-165). New York: Behavioral Publications.
- Lawton, M. P. (1975). The Philadelphia geriatric center morale scale: A revision. *Journal of Gerontology*, 30(1), 85-89.
- Lucas, R. E., & Lawless, N. M. (2013). Does life seem better on a sunny day? Examining the association between daily weather conditions and life satisfaction judgments. *Journal of Personality and Social Psychology*, 104(5), 872.
- Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, 46(2), 137-155.
- Neugarten, B. L., Havighurst, R. J., & Tobin, S. S. (1961). The measurement of life satisfaction. *Journal of Gerontology*, 16, 134-143.
- Pavot, W. (2008). The assessment of subjective well-being: Successes and shortfalls. In M. Eid & R. J. Larsen, (Eds.) *The science of subjective well-being* (pp. 124-140). New York: Guilford Press.
- Pavot, W., & Diener, E. (1993a). The affective and cognitive context of self-reported measures of subjective well-being. *Social Indicators Research*, 28, 1-20.
- Pavot, W. & Diener, E. (1993b). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5, 164-172.
- Pavot, W., & Diener, E. (2008). The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, 3, 137-152.
- Pavot, W., Diener, E., Colvin, C. R., & Sandvik, E. (1991). Further validation of the Satisfaction With Life Scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*, 57, 149-161.
- Pavot, W., Diener, E., & Suh, E. (1998). The Temporal Satisfaction With Life Scale. *Journal of Personality Assessment*, 70, 340-354.
- Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385-401.
- Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non self-report measures. *Journal of Personality*, 61(3), 317-342.
- Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4(3), 219.
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, 67(6), 1063.
- Schneider, L., & Schimmack, U. (2010). Examining sources of self-informant agreement in life-satisfaction judgments. *Journal of Research in Personality*, 44(2), 207-212.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.
- Scollon, C. N., Diener, E., Oishi, S., & Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of Cross-Cultural Psychology*, 35, 304-326.
- Seligson, J. L., Huebner, E. S., & Valois, R. F. (2003). Preliminary validation of the brief multidimensional students' life satisfaction scale (BMSLSS). *Social Indicators Research*, 61(2), 121-145.
- Strack, F., Martin, L. L., & Schwarz, N. (1988). Priming and communications: Social determinants of

information use in judgments of life satisfaction. *European Journal of Social Psychology*, 18, 429-442.

Su, R., Tay, L., & Diener, E. (2014). The development and validation of the Comprehensive Inventory of Thriving (CIT) and the Brief Inventory of Thriving (BIT). *Applied psychology: Health and Well-Being*, 6(3), 251-279.

Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology*, 74 (2), 482-493.

Tay, L., Chan, D., & Diener, E. (2014). The metrics of societal happiness. *Social Indicators Research*, 117(2), 577.

Veenhoven, R., (2017) *World Database of Happiness*, Erasmus University Rotterdam, The Netherlands. Accessed on June 25, 2017 at: <http://worlddatabaseofhappiness.eur.nl>

Vitterso, J., Roysamb, E., & Diener, E. (2002). The concept of life satisfaction across cultures: Exploring its diverse meaning and relation to economic wealth. In E. Gullone, R. Cummins (Eds.), *Social Indicators Research Book Series: The universality of quality of life constructs*. Dordrecht: Kluwer Academic.

Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form. University of Iowa: Iowa Research Online.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.

World Values Study Group. *World Values Survey, 1981-1984 and 1990-1993* (ICPSR version) [Electronic data file]. (1994). Ann Arbor, MI: Institute for Social Research [Producer and Distributor].

Yap, S. C. Y., Wortman, J., Anusic, I., Baker, S. G., Scherer, L. D., Donnellan, M. B., & Lucas, R. E. (in press). The effect of mood on judgments of subjective well-being: Nine tests of the judgment model. *Journal of Personality and Social Psychology*.



© 2018 Ed Diener. Copyright Creative Commons: Attribution, noncommercial, no derivatives