3

Assessments of Societal Subjective Well-Being

Ten Methodological Issues for Consideration

Louis Tay, Andrew T. Jebb, and Victoria S. Scotney

Abstract

This chapter examines 10 methodological issues when assessing and analyzing societal well-being using self-reports. First, there are unit-of-analysis issues: deciding the appropriate level of analysis, accounting for individual-level score variability in societal-level scores, testing isomorphism across levels, and finding ways of aggregating and accounting for score variability. Second, there are comparability issues: researchers have sought to homogenize well-being scales with different response scales or use translated measures to compare across nations. Furthermore, there is the concern of whether well-being measures can capture the full range of well-being (both positive and negative aspects). The final set of issues are prediction issues: well-being measures may be more sensitive to negative than positive events/experiences, societal well-being may not always be linearly related to variables of interest, and domain-specific measures may be more sensitive than general measures of well-being, especially when tracking specific changes in well-being or comparing subgroups.

The topic of well-being has been one of the perennial concerns in human history, although its definition and manifestations across time and cultures have varied (McMahon, 2006). In this day and age, individuals and governments continue to care deeply about well-being, and we have come to a consensus that, despite the different philosophical, cultural, and historical traditions concerning it, different psychological dimensions of well-being can be

Louis Tay, Andrew T. Jebb, and Victoria S. Scotney, *Assessments of Societal Subjective Well-Being* In: *Measuring Well-Being*. Edited by: Matthew T. Lee, Laura D. Kubzansky, and Tyler J. VanderWeele, Oxford University Press (2021).
 DOI: 10.1093/oso/9780197512531.003.0004

empirically assessed (Diener, 1994). These scientific advances have prompted the use of well-being assessments as part of national accounting among governments and intergovernmental agencies (Diener, Oishi, & Tay, 2018). For example, in 2010, the former Prime Minister of the United Kingdom, David Cameron, advocated for well-being as the key indicator of progress rather than the gross domestic product (GDP). In 2011, the United Nations General Assembly adopted historic Resolution 65/309 that recognized the limitations of GDP as an index of societal progress and encouraged member nations to develop new indicators for measuring happiness and well-being.

While well-being can be assessed in any given society, there remain multiple issues that can limit its use and usefulness. For example, Paul Allin (Chapter 2, in this volume) notes that even though well-being is being tracked, such information is not readily used in policy-making. In this chapter, we focus on methodological issues that should be highlighted and considered when examining societal levels of subjective well-being. We believe that awareness of these methodological issues will better enable us to address potential limitations in its assessment and analysis. We use the term "subjective well-being" to refer to *self-reported* well-being given that such methods have a long history (Schwarz, 1999) and continue to be the primary way for assessing well-being. Other newer data science methods such as internet searches (e.g., Ford, Jebb, Tay, & Diener, 2018) and deriving assessments of well-being from social media data (e.g., Schwartz et al., 2013) will not be directly discussed here due to space limitations, although many of the issues raised are also applicable to such approaches. Furthermore, we do not cover psychological processes, such as retrospective biases or survey techniques such as scale response labeling (e.g., Kahneman, 1999; Schwarz & Strack, 1999) as these are well-known and have been discussed elsewhere (e.g., Tay, Chan, & Diener, 2014). In this regard, we seek to raise methodological issues that are less studied. For each issue, we provide recommendations for addressing them and make suggestions for future research.

Unit-of-Analysis Issues

Much of the research on the reliability and validity of subjective well-being indicators has emerged from the psychological literature (e.g., Diener, 1984; Ryff & Keyes, 1995). The psychological focus and a methodological reliance on individual respondents to provide self-reported subjective well-being scores has naturally led to an examination of the psychometric properties of well-being measures at the *individual* level of analysis. Calculating the properties of these well-being measures does not take into consideration group-level units (e.g., communities, countries). For example, an internal consistency reliability of 0.82 and a 2-month test-retest reliability of 0.82 for the Satisfaction with Life Scale (SWLS) is calculated using individual-level scores (Diener, Emmons, Larsen, & Griffin, 1985). Yet, in seeking to assess subjective well-being at a societal level, we need to use a different level as the primary unit of analysis to understand reliability and factorial validity and even to aggregate different distributions of individual subjective well-being scores. These are all unit-of-analysis issues.

Issue 1: The Importance of Different Societal Levels (Not Just the Nation)

When considering different levels of societal subjective well-being, the first units that come to mind are usually countries. One's national identity is a salient part of personal psychological identity, making the nation level a very important level to address when considering well-being (e.g., Morrison, Tay, & Diener, 2011). However, at the same time, there are other important societal levels that can be considered. Imagine an individual living in a borough of New York City. This person will exist within and relate to many important societal levels beneath the nation. A particularly important one is the *community level*. Steps above the community level for this person will be the greater *metropolitan level* (e.g., the greater New York City area), the *state and province level* (New York), and the *regional* level (Eastern United States). At any one of these levels, subjective well-being assessments can be made by aggregating individual ratings of well-being, which may reveal observable differences between units existing at that level. These differences may reveal important events or circumstances at that level that are occurring and impacting individuals collectively.

In addition to societal levels below the nation, there are also societal levels above it. Examples occur when countries are aggregated into world regions, as seen in the *CIA Factbook*, and, of course, the broadest societal level: human civilization as a whole. As one advances to higher levels, accurate assessment is usually more time-consuming because of the increased scope, but the benefit is that these measurements provide a greater overall summary of human well-being. Thus, well-being assessments at lower and higher societal levels complement one another: lower levels have a smaller span that may be more directly relevant to the lives of the member individuals, whereas higher level assessments provide broader summaries due to their larger scope. This means that although countries are psychologically important, societal well-being can be conceived at many other levels of analysis, and any of these may be fruitful and meaningful to measure. For example, at the community level, one may be interested in the characteristics of the neighborhood that could impact well-being (Luttmer, 2005); however, at the regional level, one may be interested in the impact of culture on well-being (Diener, Oishi, & Lucas, 2003).

There are two senses of the term "societies": one based on individuals living in proximity to one another; the other based on individuals who share a common purpose. Therefore, apart from societal levels that are distinguished or based on regions, another societal level is the *organizational level*, where the interest is organizational units such as schools, companies, or institutions. In this regard, individuals are sampled across multiple organizations to make comparisons, such as the well-being of schools within the United States. At the organizational level, members in the units of interest can be situated within a specific region. However, these members can also be situated across different regions. For instance, organizations such as multinational companies have organizational members than span across multiple nations. Therefore they may not necessarily be nested, unlike levels based on regions.

Methodologically, delineating these different types of societal levels can also be helpful in deciding the types of analysis to use. When regions are used as the unit of analysis, there is a clear nested structure: communities are nested within states, states are nested within nations, and nations nested within broader regions. Critically, we need to recognize that these effects across multiple levels occur concomitantly to potentially influence subjective well-being. It is important to consider the use of statistical techniques that account for the nested structure of the data, such as multilevel modeling (Snijders & Bosker, 1999) if we are to parse the extent that these different levels exert effects on well-being. At the organizational level, members of these groups may have multiple shared groupings (e.g., multinational companies have members in different nations but share a common company). However, statistical techniques such as (generalized) linear mixed modeling (Berridge & Crouchley, 2011) can similarly be considered to account for these different types of groupings in the data.

Recommendation: When examining societal-level well-being, researchers should be very mindful of the different collective levels that exist and can be assessed. They should avoid thinking that the nation level is the only important societal level. Many important assessments can be seen at lower (e.g., communities, states/provinces) or higher levels (e.g., world regions). Furthermore, given that multiple levels exist, analysis of societal subjective well-being data at a specific level (e.g., community level) may need to account for other possible levels (e.g., nation level) through appropriate statistical modeling.

Issue 2: Variance Attributable to a Societal Level and Reliability

When we are using individual-level self-reports aggregated to a societal level, it is important to ask whether well-being measures discriminate between societal units of interest (e.g., countries, communities). One way to assess this is by examining the variability, or variance, of aggregated (to a societal level) well-being scores. Intuitively, if there is substantial variability across these aggregated scores, then one can tell which societies have higher well-being and which have lower well-being. If there is minimal variability or variance in aggregated scores, then the well-being measure is less useful in differentiating societies.

Recommendation: Where possible, we should consider examining the absolute variability (e.g., standard deviation) of societal-level scores.

For simplicity, let us consider two primary levels—individuals and nations. A more appropriate way of examining this issue is not merely assessing the absolute variability of national scores but also assessing how much of the total variability of a measure is due to the individual versus national level. This represents the extent we can expect national-level factors to influence a specific measure of well-being and is substantively important for researchers seeking to understand relative influences of national-level factors versus individual-level factors. For example, researchers often calculate *intraclass correlations* (ICCs), which are defined as the between-group variability (i.e., national-level variability) divided by total variability (i.e., national-level variability + individual-level variability). In one study, different well-being measures on a sample of 123 nations showed that the ICC (specifically ICC(1)) for life satisfaction was 0.24, meaning that 24% of the variance of life satisfaction was attributable to the national level (Tay & Diener, 2011). By contrast, ICCs for positive and negative emotions were 0.06 and 0.04, respectively. Because the ICC represents the proportion of variance attributable to the national level, we expect that the measure of life satisfaction may be relatively more sensitive to national-level factors (e.g., GDP per capita) than positive and negative emotion measures. Notably, the ICC index is also the basis for computing reliability of individual self-report measures aggregated to the national level. Having greater reliability at the national level means that we are more likely to be sensitive to picking up effects at the national level (Bliese, 1998).

> Recommendation: Where possible, we should consider examining the proportion of variability attributable to the societal level(s) of interest (or higher level of aggregation), usually quantified using the ICC(1). Similarly, where possible, we should calculate the reliability of societal-level subjective well-being scores.

Issue 3: Isomorphism and Factorial Validity

Another significant point to consider is whether the factorial structure of a measure at the individual level holds when scores are aggregated. We do not necessarily know if the same factor structure will hold at a societal level. For instance, a measure of well-being may be composed of several latent variables, but what if there is a different number of factors when the scores are aggregated to a higher societal level? This would mean that distinguishing these different dimensions may be only viable at the individual level but not at a societal level; that is, that the concept being measured is fundamentally different. If this is the case, the score averages using individual-level factorial structures cannot be compared across different societies. Furthermore, we may not be able to appropriately label the different dimensions of subjective well-being (i.e., life evaluations, positive emotions, negative emotions) given that we do not know if they are similarly distinct dimensions at a societal level. We often assume that the factor structure remains preserved when making score aggregations, but we do not often directly examine it. This issue has been recognized in other domains of research such as the isomorphism of values between individual and societal levels. Past research on values isomorphism has found high similarity between the two levels but that they are not strictly isomorphic (Fischer, Vauclair, Fontaine, & Schwartz, 2010).

Examining whether the factor structure remains constant across levels is called testing for *psychometric isomorphism* (Tay, Woo, & Vermunt, 2014). Psychometric isomorphism is essentially a kind of measurement invariance. Classic measurement invariance seeks to show that a measure has the same factor structure across groups at the same level of analysis (e.g., between men and women or respondents from different countries) to show that the same construct is being measured. Testing for psychometric isomorphism does the same thing but with regard to different levels of analysis. One can conveniently think of classic measurement invariance as testing the measure's factor structure "horizontally," whereas psychometric isomorphism tests it "vertically." The reader should note that this only needs to be done when scores are aggregated using a method like taking the mean (not when aggregating to get a measure of variability like the variance; Jebb, Tay, Ng, & Woo, 2019).

> Recommendation: Where possible, tests of psychometric isomorphism should be conducted between the individual level and a societal level that the scores are being aggregated to (Tay, Woo, & Vermunt, 2014). This is done to ensure that the factor structures hold across levels so that the same well-being construct(s) can be said to be measured.

Issue 4: Limitations in Measures of Central Tendency

Researchers and policy-makers often rely on the statistical mean values of individual scores to represent societal-level scores. This may be done out of habit or convenience because there are other statistical parameters that can be used to assess the center of the distribution, such as the mode or median. More critically, all of these are simply measures of the distribution's *central tendency* and do not show how the scores within a particular level of society are distributed.

In addition to central tendency, there are measures of well-being distributions that could be examined. This is substantively informative to consider, as what matters is not only the average happiness levels for a society, but also the degree of *dispersion* (or spread) in well-being scores. Whether there is a large or a small spread can indicate important things about that society, such as an unequal distribution of resources or inequality among societal members. Such distributions are critical for researchers and policymakers to understand because they can shed light on who may have substantially lower and higher levels of well-being and what types of factors may be at play. Indeed, Veenhoven (1990) notes that there are different types of inequalities that may exist in societies (e.g., income, power, prestige, etc.) and that statistics of the well-being distribution, rather than central tendency, may better reveal these. Specifically, in addition to central tendency, other measures that can be used are dispersion measures (e.g., standard deviation, range, variance) or measures of skew (for examples, see Jebb, Tay, Ng, & Woo, 2019). Like the mean, measures of dispersion can serve as a relevant outcome or predictor in statistical modeling. Considering the societal-level of the nation, past research shows that the Gini coefficient calculated on life satisfaction scores (which is a measure of dispersion in the distribution) predicts average levels of enjoyment, anger, sadness, and stress over and above average annual household income (Diener & Tay, 2015). Moreover, Ovaska and Takashima (2014) have found income inequality and health inequality to be related to higher levels of life satisfaction dispersions within countries.

> Recommendation: Researchers examining societal-level well-being should recognize the limitations of central tendency measures for describing the well-being distribution and the importance of other statistics, such as dispersion and skew measures. Visualizations of the full distribution (e.g., histograms, kernel densities) can be an effective way to see or communicate the full picture of well-being at that level of society.

Comparability Issues

The next set of issues are collectively organized under comparability issues, where we discuss how past work has sought to create comparable well-being scores across societies—typically nations. We examine some of these methods and provide recommendations. Furthermore, we also consider whether subjective well-being measures capture the full range of well-being (i.e., equivalence in assessing positive and negative aspect of well-being in a measure).

Issue 5: Scale Homogenization via Linear Stretch

In survey research, it is common to see different measures for the same well-being construct. Sometimes researchers simply develop or use different measures because they are working independently (e.g., there are many slightly different life satisfaction measures). At other times, this occurs with archival data from different countries that have assessed well-being at different time periods in that nation. In any case, because scales are not always standardized over research studies, it is hard to know whether these scores can be compared. Scales of the same construct might have different numbers of response options (e.g., 1–5 vs. 1–10 response options) or types of response options (e.g., Not at all/A little bit/Somewhat/Very much/extremely vs. strongly agree/Agree/Disagree/Strongly disagree). Are two 1–5 scales with different response options perfectly comparable?

The different types and wording of the items notwithstanding, researchers have sought to use transformations to create a common metric to compare different measures across countries and over time. In other words, researchers seek to conduct *scale homogenization* (de Jonge, Veenhoven, & Arends, 2014) to either "stretch out" scores to a common standard (e.g., 1–5 response scale to a 1–10 response scale) or to "compress" them (e.g., 1–11 response scale to a 1–10 response scale). One conventional scale homogenization method has been the *linear stretch* method, which takes the lowest scale response option (e.g., 1 on a 1–5 response scale) and projects it onto the lowest number on a common scale (e.g., 0 on a 0–10 response scale) and the highest response option (e.g., 5 on a 1–5 response scale) and projects it onto the highest number on the common scale (e.g., 10 on a 0–10 response scale). All intermediate scale response options are then transformed to equally distanced scores (e.g., 2 on a 1–5 response scale projected as 2.5 on a 0–10 response scale; de Jonge et al., 2014).

Past research has shown that the linear stretch method for scale homogenization does not lead to equivalence. Using the World Database of Happiness of 67 nations between 1945 and 2013, it was found that when scales were equivalized to a 0–10 scale, those with fewer scale response options on the original scale (e.g., 3 response options vs. 7) had lower rescaled scores (Batz, Parrigon, & Tay, 2016). This trend held for both life satisfaction and other happiness scales. In other words, the linear stretch method may *artifactually* score nations lower if they used self-reported well-being scales that had few response options. In addition to distorting raw scores, this can also affect substantive relationships. In the same study, GDP per capita predicting happiness had linear regression coefficients that were significantly larger when not accounting for the number of scale response options ($\beta = 0.65$, 95% confidence interval [CI] [0.49, 0.81]) as compared to accounting for the scale response options ($\beta = 0.33$, 95% CI [0.22, 0.44]; Batz et al., 2016).

Instead of the conventional linear stretch method of scale homogenization, past work has shown that other more sophisticated methods may be more appropriate, although with limitations (de Jonge et al., 2014). Our suggestion is that researchers interested in societal well-being should strive to use scales that are homogenous in the first place rather than rely on mathematical transformations.

Recommendation: For comparisons of societal-level well-being, researchers must be aware that when different scales are being used, especially if the number of response options differs, there may not be equivalence. Ideally, research should rely on common, equivalently worded scales, scale responses, and scale response wording for accurate comparisons.

Issue 6: Measurement Equivalence

Even with the use of equivalently worded measures—both in item content and scale responses—there can be differences in the language(s) spoken and understood by respondents in a society that may lead to nonequivalence in measurement. In other words, language differences can lead to potential measurement bias in well-being scales; *measurement bias* means that respondents in a society are systematically scored higher or lower than their "true" level (Tay, Meade, & Cao, 2015). Therefore, we may observe score differences on well-being scales between societies not because of "true" construct differences but because the well-being scales are not equivalent. Aside from language, there are other reasons for measurement bias that include differences in culture (even with the same language; e.g., culture of modesty), differences in response context (e.g., phone poll vs. face to face), and other factors (Robert, Lee, & Chan, 2006).

Therefore, it is important to not merely use equivalently worded measures and have appropriate translations into different languages, but to also empirically determine if there is measurement equivalence across societal units of interest (e.g., communities, countries, organizations). There are well-established statistical procedures for assessing measurement equivalence (Tay et al., 2015; Vandenberg & Lance, 2000). A key requirement is that the scale have multiple items and not be made up of a single item. Multi-item scales make it statistically possible to disentangle potential measurement bias from true differences, whereas a single item cannot. Moreover, with multi-item measures we are able to ascertain if there are specific items that may be causing any nonequivalence. For example, it has been shown that even within the United States, a negatively scored item for negative affect (i.e., "full of life") was not equivalent between Hispanics/Latinos and non-Hispanic whites (Kim, Wang, & Sellbom, 2020). Identifying these nonequivalent items will allow researchers to later exclude them from the analysis. This can only be done if multi-item measures are used.

For example, our research group examined the measurement equivalence of the Comprehensive Inventory of Thriving (CIT; Su, Tay, & Diener, 2014), which assesses a large number of dimensions of well-being (e.g., relationship, engagement, mastery, autonomy, meaning, optimism, subjective well-being). Measurement equivalence was assessed at the societal level of nations: United States, Argentina, Australia, China, Germany, India, Mexico, Russia, Singapore, Spain, and Turkey (Wiese, Tay, Su, & Diener, 2018). It was found that there was measurement equivalence of the CIT measure across the different countries but not in Argentina, Mexico, and China. Specifically, it was found that the dimension of "engagement" did not fit well across these nations. This finding can serve to prompt further research on whether the translation of "engagement" may be interpreted differently in these countries. It can also guard against reading too much into national differences on the "engagement" dimension.

> Recommendation: Where possible, researchers should use multi-item well-being measures and assess their measurement equivalence when different societies are examined. This is not possible with a single-item measure. Additionally, it can be beneficial to use scales that have already been examined and validated for measurement equivalence across nations as this can help forestall potential problems in nonequivalence prior to data collection.

Issue 7: Equivalence in Well-Being Poles

One issue that has been more recently considered in the literature is whether our measures adequately assess the full range of well-being (i.e., from suffering to flourishing; Tay & Jebb, 2018). While positive aspects of well-being have long been considered and valued, positive psychology (Seligman & Csikszentmihalyi, 2000) has focused measurement efforts to go explicitly beyond the negative aspects (e.g., depression, stress, suicide). A comprehensive assessment of societal subjective well-being should consider whether our administered measures equivalently assess both positive and negative aspects.

In our work on continuum specification, we discuss the importance of defining and operationalizing *construct continua* (Tay & Jebb, 2018). In its application to societal subjective well-being, we need to determine whether we are adequately capturing one end of well-being to the lack of inclusion of the other. We also need to be clear about the nature of the gradations on the continuum (i.e., the quality that separates high from low scores; e.g., intensity of positive feelings; frequency of positive feelings). Continuum specification enables greater measurement validity, where we explicitly define the full span of a concept and are confident that our scale content and scale response options operationally assess this full span. For example, in claims that most people are happy (Diener & Diener, 1996), it is important to consider if the well-being scale administered adequately captures the full span in the purported well-being poles.

Continuum specification also exists at a broader conceptual analysis beyond a single self-report scale. In this regard, the continuum does not refer to the continuum of any given single construct (e.g., meaning, life satisfaction, social support). Rather, the continuum refers to the full range of the well-being concept (i.e., from suffering to flourishing). Different types of measures that index ill-being and well-being are required in the suite of societal subjective well-being measures. This may include pairings such as loneliness and social connection, positive feelings and negative feelings, depression and awe, distress and eustress, optimism and pessimism.

> Recommendation: Researchers should be mindful of all the degrees and the overall span that well-being can take. If a full measure of well-being is desired, one must ensure that the full range of well-being is encompassed or is at least included in the suite of measures that are used.

Prediction Issues

The initial set of issues was concerned with the assessment of societal well-being itself. This final set of issues is concerned with how these assessments of societal well-being are typically used in research and analysis which form the basis of scientific conclusions and policy recommendations. If societal well-being is measured, one interest will simply be in the levels that are observed. This information is important because it gives a direct summary about the well-being of a particular level of society. However, researchers will often want to go beyond examining just these differences in levels. They will often want to also examine what predicts well-being (e.g., GDP per capita, employment) and what is predicted by well-being (e.g., physical health, different social attitudes; Diener et al., 2018). We label these as *prediction issues*. In some cases, prediction will be important because one is trying to establish a causal relationship, and correlation is a necessary condition for causation. In others, prediction itself will be the goal; a variable might be able to predict societal well-being even though it is not a direct cause, and the ability to anticipate future well-being levels can be important in its own right (see Shmueli, 2010, for distinctions between causal and predictive modeling).

In this section, we touch on several issues related specifically to building statistical models that include societal well-being. However, because statistical modeling is an extensive topic, the reader should note that there are many more issues that we do not touch on here, such as consideration of interaction effects, appropriate use of control variables, and fulfilling the assumptions of these models (e.g., homogeneity of variance). These more general issues are important and are discussed in many other resources (e.g., Gelman & Hill, 2007).

Issue 8: Bad May Be Stronger Than Good

One important consideration in assessing the predictors of well-being is that negative and positive events have an asymmetrical relationship to well-being. Specifically, negative events are known to generally be more strongly related to well-being than are positive events (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). For instance, when local businesses or factories close, people are more emotionally affected than when they are preserved or added to. Empirically, pivotal research in behavioral economics has shown that people are more sensitive to losses than gains (i.e., loss aversion; Kahneman & Tversky, 1979). In psychology, there is evidence that negative events (e.g., divorce, job loss, loss of spouse) tend to exert stronger and longer lasting effects on subjective well-being than are positive events (e.g., marriage, lottery wins;

Tay & Kuykendall, 2013). These effects have also recently been demonstrated at higher societal levels; a recent study by De Neve et al. (2018) showed that negative economic change is more strongly related to national subjective well-being than is positive economic change. This negativity bias has been observed even in early development (Vaish, Grossmann, & Woodward, 2008). When positive events happen, there is less of a reason to feel strong emotions because events are unfolding as they "ought to." However, when negative events occur, strong emotions are needed to draw attention to the problems and correct them. Indeed, many theorists agree that this is a standard purpose of emotion: to direct our attention and motivate us to deal with life events (Izard, 2010).

These considerations are useful when constructing a statistical model where well-being is conceived as an outcome of societal-level indicators. From this psychological theory, we think it is likely that societal well-being is better predicted by negative indicators. Knowing this provides guidance in constructing better models of societal well-being and helps avoid looking for important predictors in the dark. For example, rather than examining very broad societal-level factors like GDP or housing rates, one might have better success looking at variables that directly quantify problems, such as the percentage of the unemployed or the percentage of those who have difficulty affording housing. That is, one interesting strategy is to look at *societal problems* rather than just broad societal factors. This approach is informed by a consistent finding that negative and positive events have an asymmetrical relationship to well-being. We can apply this to different societal levels by examining specific problems at that level (e.g., community- or state-level problems).

> Recommendation: When constructing models of societal-level well-being, researchers should recognize that positive and negative events may not be equally related to well-being. Using this psychological theory, researchers may be able to more efficiently identify the key predictors, which might be problems that exist at that societal level.

To clarify, we are not stating that bad will *always* be stronger than good or that protective or positive factors will have nugatory effects. Rather, when we are seeking to model predictors of well-being from a large number of variables at a specific societal level, it is worth considering negative events or conditions. Certainly, researchers may be interested in comparing different types of positive conditions and their differential effects on well-being or the different types of positive well-being dimensions and their differential effects on outcomes (e.g., Chapters 4 and 5, both in this volume). Even when seeking to assess positive conditions, one can assess it alongside negative conditions or variables, as when comparing between positive and negative economic changes on societal well-being (De Neve et al., 2018). This also extends to modeling outcomes of well-being where positive and negative emotions have been found to have independent effects on future physical health (Wiese, Chen, Tay, Friedman, & Rector, 2018).

Issue 9: Curvilinear Effects

Linear models are prevalent in the social sciences. Part of this is due to the fact that they are both conceptually and mathematically simple. Conceptually, *linearity* means that the slope between the variables remains constant, and mathematically, any linear relationship can be represented by simple arithmetic terms (e.g., multiplication and addition). However, the popularity of linear models is not just due to convenience, but also because they are often accurate. Linearity is often a useful approximation for real-life processes (Cowpertwait & Metcalfe, 2009). For example, it is natural to think of well-being as increasing linearly with things like amount of leisure time or financial security. However, because well-being cannot increase infinitely (Diener, Lucas, & Scollon, 2006), it is likely that it will have many associations that are not fully linear. With the examples of leisure time and financial security, one can easily imagine that, after an individual has enough leisure time or financial security, then more of either of these does not increase well-being. Indeed, recent research has given evidence that the relationship between subjective well-being and income is linear but only up to a point. When income is high enough to satisfy basic needs and higher order desires, then the relationship becomes flat (Jebb, Tay, Diener, & Oishi, 2018). Instead of linearity, this is an example of a *monotonic* relationship, which is simply any relationship between variables where the slope maintains the same sign (but is not necessarily linear because the magnitude of the slope can change). In addition to monotonicity, there are also *curved relationships* where the sign of the slope *does* change (from positive to negative or vice versa). One can easily imagine a scenario where all of a person's time is *leisure*, which can lead to feelings of a lack of meaning and negatively impact well-being. One empirical example can be found in the individual-level association between well-being and age. Substantial research has shown that the association between well-being and age has a slight curve such that, prior to middle age, the slope is negative, but after middle age, the slope is positive (Blanchflower & Oswald, 2017; López Ulloa, Møller, & Sousa-Poza, 2013). At a societal level, GDP per capita does not increase linearly with national subjective well-being (e.g., Inglehart, Foa, Peterson, & Welzel, 2008).

The key point we hope to make is that just because linearity often holds at some levels of the variables, researchers should not be lured into thinking that the *entire* relationship is necessarily linear. In this regard, we may want to consider some alternatives to the standard linear model, such as polynomial models (Cohen, Cohen, West, & Aiken, 2003), regression splines (Harrell, 2015), or nonlinear models (Motulsky & Ransnas, 1987) when modeling societal-level well-being.

> Recommendation: Carefully consider whether a relation may be linear when analyzing societal well-being. One may need to utilize other models aside from standard linear regression to accurately estimate and understand the predictors and outcomes of societal well-being.

Issue 10: Level of Specificity

Because well-being can be considered to encompass many aspects of life, assessments of well-being will vary in whether they are *broad* (e.g., satisfaction with life in general, negative emotions) or *specific* (e.g., satisfaction with work, the presence of worry). This is called the *level of specificity* of the measure, and it is an important factor to consider when assessing or modeling societal well-being. First, in terms of assessment, a greater level of specificity helps respondents focus on a specific domain of well-being in question (e.g., Cummins, 2005; Oishi & Diener, 2001). This makes that domain the source of their well-being, making it arguably more sensitive to that source. Second, when including well-being as a variable in statistical models, specificity is important to consider because, in general, specific outcomes tend to be predicted better by specific predictors, whereas broader outcomes are better predicted by broader predictors (Cronbach & Gleser, 1965). When the outcome is broad, many aspects are measured as part of it, and if this content is not matched to the content in a predictor, the correlation will be driven down. Conversely, if the outcome is specific and the predictor is broad, the association will be reduced by the content in the predictor that is simply irrelevant to the outcome. For instance, an overall measure of physical health includes many components (e.g., illnesses, quality of diet), and a broad predictor like an overall happiness score may be a good fit because it is also a function of many things (e.g., satisfaction with one's job, social life, and leisure activities). However, as the outcome becomes narrower in scope, more and more of a broad predictor will become irrelevant. Thus, predicting a specific aspect of physical health, like exercise habits, will have a lower correlation to overall happiness. In this case, a better predictor would be a more specific measure, such as vitality.

The idea that associations are maximized when outcomes and predictors are matched in their level of specificity is referred to as the *bandwidth-fidelity* distinction (Cronbach & Gleser, 1965). This is just an observation to consider when planning a research study, and it should not determine what the researcher does. Predictors included in a statistical model can be broad even if the outcome is narrow (and vice versa); it depends on the context and aims of the research (e.g., maybe the researchers is genuinely interested in how overall well-being relates to a specific narrow variable). However, we note this because, to our knowledge, this is not an issue often considered explicitly in well-being research. Our research group sought to examine this issue by comparing the sensitivity of life satisfaction (very broad) and job satisfaction (more specific) to gender inequality and found preliminary evidence that job satisfaction may be more sensitive to gender inequality than overall life satisfaction (Batz-Barbarich, Tay, Kuykendall, & Cheung, 2018). This suggests, when looking at specific factors that might reduce (or promote) well-being, that one should consider using more specific well-being measures.

> Recommendation: For a given research question or topic, one should consider the level of specificity at play. If the outcome or predictor is narrow, this means its content will also be narrow. Generally, broad outcomes tend to have stronger associations with broad predictors, and the same is true for narrower outcomes and predictors. When investigating what best predicts narrower aspects of well-being, researchers may want to consider narrower predictors.

Conclusion

The assessment of societal well-being is a growing and active area, important not only to researchers, but to government leaders and policy-makers as well. We have summarized some of the key issues for consideration and proposed specific recommendations for addressing—or at least being mindful of—these issues. We presented several suggestions to consider when assessing well-being when societies are the unit of interest. For example, researchers may want to consider the use of measures of dispersion, rather than central tendency, in order to address the range of well-being experienced within a society (e.g., What is the significance of a country where most people have similar levels of well-being [low dispersion] versus a country where there are many people experiencing well-being extremes [high dispersion]?). Furthermore, we discussed issues in comparing well-being across societies where different measurement scales are used or when items in a scale do not have the same interpretation by people of different cultures, language, or other demographics. Addressing these issues is foundational to making comparisons between societies that are not confounded by measurement artifacts. Finally, we discussed issues in prediction of and by societal well-being that include the valence of well-being measures, the linearity of relations with well-being, and the specificity of well-being measures. We hope the issues we presented encourage researchers to expand their thinking about the measurement and analysis of societal subjective well-being. We believe that the consideration of these issues could help develop new research ideas and clarify the phenomena of interest. In this regard, we hope that this chapter can serve to highlight areas that require more active research.

About the Authors

Louis Tay is Associate Professor of Industrial-Organizational Psychology at Purdue University. His research interests are in well-being, methodology, and measurement. He is the co-editor of the books *Handbook of Well-Being* (DEF Publishers) and *Big Data in Psychological Research* (APA Books). He serves as an associate editor at *Organizational Research Methods*.

Andrew T. Jebb is a research scientist at the University of Oklahoma's Institute for the Study of Human Flourishing. He is a graduate of Purdue University's Department of Psychological Sciences (in industrial–organizational psychology). He is interested in happiness and well-being, the employee–organization relationship, and psychological measurement and statistics.

Victoria S. Scotney is a graduate student in industrial–organizational psychology at Purdue University. She has published in outlets such as *Psychological Bulletin* and *Frontiers*

in Psychology. Her current research interests include prosocial giving and receiving, well-being, research methods, and measurement.

Author Note

This work was supported in part by a grant from the John Templeton Foundation and by the Lee Kum Sheung Center for Health and Happiness. The views expressed in this chapter represent the perspectives of the authors and do not reflect the opinions or endorsement of any organization. We have no known conflict of interest to disclose.
Correspondence concerning this chapter should be directed to Louis Tay, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, Indiana, 47907 (stay@purdue.edu).

References

Batz, C., Parrigon, S., & Tay, L. (2016). The impact of scale transformations on national subjective well-being scores. *Social Indicators Research*, *129*, 13–27.

Batz-Barbarich, C., Tay, L., Kuykendall, L., & Cheung, H. K. (2018). A meta-analysis of gender differences in subjective well-being: Estimating effect sizes and associations with gender inequality. *Psychological Science*, *29*(9), 1491–1503.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370.

Berridge, D. M., & Crouchley, R. (2011). *Multivariate generalized linear mixed models using R*. Cleveland, OH: CRC Press.

Blanchflower, D. G., & Oswald, A. J. (2017). Do humans suffer a psychological low in midlife? Two approaches (with and without controls) in seven data sets. Bonn: Institute of Labor Economics (IZA). Https://www.econstor.eu/bitstream/10419/170942/1/dp10958.pdf

Bliese, P. D. (1998). Group size, ICC values, and group-level correlations: A simulation. *Organizational Research Methods*, *1*(4), 355–373.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cowpertwait, P. S., & Metcalfe, A. (2009). *Introductory time series with R*. New York: Springer-Verlag.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Champaign: University of Illinois Press.

Cummins, R. A. (2005). The domains of life satisfaction: An attempt to order chaos. In A. C. Michalos (Ed.), *Citation classics from social indicators research: The most cited articles edited and introduced by Alex C. Michalos* (pp. 559–584). Dordrecht: Springer Netherlands.

de Jonge, T., Veenhoven, R., & Arends, L. (2014). Homogenizing responses to different survey questions on the same topic: Proposal of a scale homogenization method using a reference distribution. *Social Indicators Research*, *117*, 275–300.

De Neve, J. E., Ward, G., De Keulenaer, F., Van Landeghem, B., Kavetsos, G., & Norton, M. I. (2018). The asymmetric experience of positive and negative economic growth: Global evidence using subjective well-being data. *Review of Economic Statistics*, *100*(2), 362–375.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*, 542–575.

Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research*, *31*, 103–157.

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological Science*, *7*, 181–185.

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., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *American Psychologist*, *61*(4), 305–314.

Diener, E., Oishi, S., & Lucas, R. E. (2003). Personality, culture and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology*, *54*, 403–425.

Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behavior*, *2*, 253–260.

Diener, E., & Tay, L. (2015). Subjective well-being and human welfare around the world as reflected in the Gallup World Poll. *International Journal of Psychology*, *50*, 135–149.

Fischer, R., Vauclair, C.-M., Fontaine, J. R. J., & Schwartz, S. H. (2010). Are individual-level and country-level value structures different? Testing Hofstede's legacy with the Schwartz value survey. *Journal of Cross-Cultural Psychology*, *41*, 135–151.

Ford, M. T., Jebb, A. T., Tay, L., & Diener, E. (2018). Internet searches for affect-related terms: An indicator of subjective well-being and predictor of health outcomes across US states and metro areas. *Applied Psychology: Health and Well-Being*, *10*(1), 3–29.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Harrell Jr, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. New York: Springer.

Inglehart, R., Foa, R., Peterson, C., & Welzel, C. (2008). Development, freedom, and rising happiness: A global perspective (1981-2007). *Perspectives on Psychological Science*, *3*, 264–285.

Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review*, *2*(4), 363–370.

Jebb, A. T., Tay, L., Diener, E., & Oishi, S. (2018). Happiness, income satiation, and turning points around the world. *Nature Human Behavior*, *2*, 33–38.

Jebb, A. T., Tay, L., Ng, V., & Woo, S. (2019). Construct validation in multilevel studies. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis* (pp. 253–278). American Psychological Association. https://doi.org/10.1037/0000115-012

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kim, G., Wang, S. Y., & Sellbom, M. (2020). Measurement Equivalence of the Subjective Well-Being Scale Among Racially/Ethnically Diverse Older Adults. *The Journals of Gerontology: Series B*, *75*(5), 1010–1017.

López Ulloa, B. F., Møller, V., & Sousa-Poza, A. (2013). How does subjective well-being evolve with age? A literature review. *Journal of Population Ageing*, *6*(3), 227–246.

Luttmer, E. F. P. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly Journal of Economics*, *120*, 963–1002.

McMahon, D. M. (2006). *Happiness: A history*. New York: Grove.

Morrison, M., Tay, L., & Diener, E. (2011). Subjective well-being and national satisfaction. *Psychological Science*, *22*, 166–171.

Motulsky, H. J., & Ransnas, L. A. (1987). Fitting curves to data using nonlinear regression: A practical and nonmathematical review. *FASEB Journal*, *1*, 365–374.

Oishi, S., & Diener, E. (2001). Re-examining the general positivity model of subjective well-being: The discrepancy between specific and global domain satisfaction. *Journal of Personality*, *69*, 641–666.

Ovaska, T., & Takashima, R. (2014). Does a rising tide lift all the boats? Explaining the national inequality of happiness. *Journal of Economic Issues*, *44*(1), 205–224.

Robert, C., Lee, W. C., & Chan, K.-Y. (2006). An empirical analysis of measurement equivalence with the INDCOL measure of individualism and collectivism: Implications for valid cross-cultural inference. *Personnel Psychology*, *59*, 65–99.

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*, 719–727.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L. A., Lucas, R. E., Agrawal, M., . . . Ungar, L. H. (2013). *Characterizing geographic variation in well-being using tweets*. Paper presented at the Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM), Boston, MA.

Schwarz, N. (1999). Self-reports: How the questions shape the answer. *American Psychologist*, *54*, 93–105.

Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York: Russell Sage Foundation.

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*, 5–14.

Shmueli, G. (2010). To explain or predict? *Statistical Science*, *25*, 289–310.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

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

Tay, L., Chan, D., & Diener, E. (2014). The metrics of societal happiness. *Social Indicators Research*, *117*, 577–600.

Tay, L., & Diener, E. (2011). Needs and subjective well-being. *Journal of Personality and Social Psychology*, *101*, 354–365.

Tay, L., & Jebb, A. T. (2018). Establishing construct continua in construct validation: The process of continuum specification. *Advances in Methods and Practices in Psychological Science*, *1*(3), 375–388.

Tay, L., & Kuykendall, L. (2013). Promoting happiness: The malleability of individual and societal-level happiness. *International Journal of Psychology*, 159–176.

Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, *18*, 3–46.

Tay, L., Woo, S. E., & Vermunt, J. K. (2014). A conceptual and methodological framework for psychometric isomorphism: Validation of multilevel construct measures. *Organizational Research Methods*, *17*, 77–106.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383–403.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70.

Veenhoven, R. (1990). Inequality in happiness: Inequality in countries compared across countries. https://mpra.ub.unimuenchen.de/11275/1/MPRA_paper_11275.pdf

Wiese, C. W., Chen, Z. J., Tay, L., Friedman, E. M., & Rector, J. L. (2018). The role of affect on physical health over time: A cross-lagged panel analysis over 20 years. *Applied Psychology: Health and Well-Being*, *11*, 202–222.

Wiese, C. W., Tay, L., Su, R., & Diener, E. (2018). Measuring thriving across nations: Examining the measurement equivalence of the Comprehensive Inventory of Thriving (CIT) and the Brief Inventory of Thriving (BIT). *Applied Psychology: Health and Well-Being*, *10*(1), 127–148.