Progress in measuring subjective well-being
Moving toward national indicators and policy evaluations

By Alan B. Krueger* and Arthur A. Stone

Progress in science requires new tools for measuring phenomena previously believed unmeasurable, as well as conceptual frameworks for interpreting such measurements. There has been much progress on both fronts in the measurement of subjective well-being (SWB), which “refers to how people experience and evaluate their lives and specific domains and activities in their lives” (1). In 2009, the Sarkozy Commission recommended adding SWB measures as supplements to existing indicators of societal progress such as gross domestic product (GDP). In light of subsequent activity by governments and international organizations, we summarize several important advances and highlight key remaining methodological challenges that must be addressed to develop a credible national indicator of SWB and to incorporate SWB into official statistics and policy decisions.

VALIDITY AND COMPARABILITY. Subjective well-being is necessarily measured by respondents’ self-reports evaluating their life and feelings (see the photo). In some fields, such reports are the standard, e.g., assessment of pain and fatigue. Although there have been long-standing concerns about the meaning of subjective reports, emerging evidence finds that self-reports are related to biological processes and health outcomes, which increases confidence in the validity of such measures. Experimental studies find that self-reports of the extent of pain are associated with changes in blood flow to brain regions known to process pain (2). Studies have shown correspondence between subjective reports of affect and experience with immunological and hormonal measures (3, 4). Mortality has been associated with low levels of SWB (5).

Nevertheless, additional evidence is needed to support the validity of between-group comparisons, e.g., when comparing SWB across countries or demographic groups. Differential reporting styles, including interpretation of questions and response scales, by groups, has the potential to yield misleading conclusions. Some rankings of countries have been questioned on these grounds.

A recent technique for evaluating and adjusting interpersonal and intergroup differences in self-reports is the “vignette approach.” Questions about the constructs under investigation are preceded by vignettes that describe an individual at some intensity of the construct, providing an explicit comparison standard for the subsequent rating. Responses can be compared and scales adjusted if systematic intergroup differences emerge (6). This has proved promising for assessing subjective health appraisals and is being applied to well-being reports, although there are potential environmental factors that may confound comparison, such as differences in the quality of health care across communities.

Interpretation and use of response scales (i.e., options for answering questions) is likely to vary according to past experiences, cultural background, genetic factors, and immediate context. Well-known response options are verbal scales (e.g., “Extremely satisfied” through “not at all satisfied”) and numeric rating scales (e.g., 0 through 10 anchored scales, with 0 indicating the absence of some feeling). On the basis of studies of individual differences in densities of taste receptors on subjects’ tongues, Bartoshuk (7) described scale elasticity, wherein self-reports with standard response scales of taste sensation were invalid. Response options had different meanings for two groups of tasters. “Supertasters” had a “stretched-out” scale, a feature lost in current response scales. The same logic may prove useful for SWB.

Another validity issue concerns the extent to which people adapt to their circumstances. There is a critical distinction in how scales are used, where extreme events can result in “recalibration” (8), as opposed to true adaptation. In the first case, changes in SWB are not due to actual differences in experience or evaluation but simply to how the scale was used. True adaptation is defined by changes in emotional experience. There is evidence that people shift their use of scales and that they can adapt to circumstances; e.g., in response to physical disabilities or winning a lottery (see (9) for a recent meta-analysis). But adaptation is not a uniform process, and some circumstances and aspects of SWB appear relatively more or less resistant to adaptation.

CONCEPTUAL FRAMEWORK. The conceptual framework for SWB remains, in our view, the most important obstacle to developing a comprehensive national indicator. Innovative new work (10) provides a framework in which individuals’ utility depends on several fundamental, nonoverlapping aspects, such as material well-being, life satisfaction, and emotional experience. Following standard economic logic, components can be aggregated on the basis of the extent to which individuals would be willing to trade an improvement in one aspect against an improvement in another, on the margin. This method is implemented by posing hypothetical choices between situations that involve different dimensions of SWB. These marginal valuations provide weights for aggregating aspects of SWB, just as prices (marginal utilities) provide appropriate weights for aggregating components of the GDP under certain assumptions.

This research is in its infancy, and objections can be raised about its reliance on revealed preference, but it provides a coher-
ent framework for aggregating dimensions of SWB on the basis of individuals’ choices. Further refinements—perhaps including actual, as opposed to hypothetical, choices—may advance the measurement of SWB.

Kahneman and Krueger (11) seek to sidestep many requirements for a comprehensive index of SWB (such as interpersonal comparability) by focusing only on experiential SWB and measuring the percentage of time that people spend in an unpleasant state, which they call the U-index. An unpleasant state is defined as an episode in which the intensity of a negative emotion is greater than the intensity of positive emotions (see the chart). They justify the U-index in part by arguing that policy-makers often care more about minimizing misery than maximizing happiness [a theme echoed in (7)]. The U-index can be constructed with the Day Reconstruction Measure, which collects time-use data together with emotional experience. The U-index is robust in that different individuals and groups can interpret scales differently, as long as they consistently apply their interpretation to positive and negative emotions. The U-index is related to, but conceptually distinct from, more traditional measures of affect (measured with momentary and diary approaches), in that the U-index emphasizes that one dominant-negative emotion can color an entire episode or day.

Until more progress is made toward developing a credible, comprehensive index of SWB, we would emphasize the importance of separately measuring the key components of SWB (e.g., satisfaction with life, positive emotional experience, meaning in life) and keeping them distinctive.

OFFICIAL MEASURES. After the Sarkozy report, the UK Office of National Statistics (ONS) initiated a program to measure SWB in its Annual Population Survey. Evaluative (“satisfaction” with life), eudaimonic (well-being or human flourishing), and hedonic (affect in everyday life) SWB were surveyed. However, the assessment consists of only four questions and lacks information on actual time use or events in people’s lives. To partly address these concerns, ONS plans more detailed surveys of SWB. The proper way to combine ONS’s measures into a comprehensive indicator is unresolved.

The Organization for Economic Co-operation and Development (OECD) published an extensive report on measuring SWB (12) to guide national statistical offices and presented a detailed critique of ongoing efforts. The OECD’s Better Life Index finesses the problem of how to aggregate different components of well-being by allowing users to set their own weights. The OECD has assembled a “high-level” commission to analyze topics that could be informed by SWB research, such as income inequality, and the United Nations launched initiatives on well-being and sustainability (e.g., the International Day of Happiness).

Earlier this year, the National Research Council (NRC) of the U.S. National Academy of Sciences issued a report on hedonic well-being and policy (1). This stressed the importance of considering both happiness and misery in SWB. It supported a broader definition of hedonic well-being (called experiential WB) that includes pain and other forms of suffering, which the panel considered important for policy purposes. The U.S. Bureau of Labor Statistics has incorporated an affective module in the American Time Use Survey (ATUS) in 2010 and 2012 to combine SWB data (happy, pain, sad, stress, tired, meaningful) with time-use information during representative periods of the day. The NRC report highlighted that, “The ATUS SWB module is practical, stable, inexpensive, and worth continuing… Not only does the ATUS SWB module support research; it also generates information to help refine SWB measures that may be considered for future additions to official statistics” (7).

A striking feature of the OECD and NRC reports is their optimism about future prospects of SWB measures. Another recent report on well-being and policy stated “we should measure wellbeing more often and do so comprehensively… This would help governments improve policies, companies raise productivity, and people live more satisfying lives” ([12] executive summary).

POLICY EVALUATION. Measures of SWB have become key outcome measures in program evaluation, often yielding deeper insights than traditional measures. For example, 10 to 15 years after the launch of the Moving to Opportunities experiment, which randomly offered some low-income families living in U.S. public housing the opportunity to move to less-disadvantaged neighborhoods, significant improvements were found for components of SWB (distress, depression, anxiety, and calmness) but not for economic (employment and earnings) and educational outcomes (14). Evaluations of the Oregon Medicaid expansion found significant improvements in subjective outcomes, including depression and self-reported health, but mixed results concerning physical health (15). Because individuals and policy-makers value subjective outcomes and because such outcomes appear to be affected by major policy interventions, measures of SWB are likely to play an increasingly important role in policy evaluation (16) and decisions.

Further advances in measurement and the conceptual framework for combining SWB measures are needed before sufficient consensus can be reached to support a comprehensive, official measure of SWB that can be compared at the national level. In the meantime, comparisons of various components of SWB, with a focus on negative emotional experiences, strike us as a reasonable agenda for national statistical agencies and researchers.

REFERENCES AND NOTES

15. F. Finkelstein et al., Oregon Health Study Group, Q. J. Econ. 127, 1057 (2012).

ACKNOWLEDGMENTS

A.B.K. was a member of the 2009 Sarkozy Commission and the ONS Forum on Measuring National Well-Being. A.A.S. chaired the National Academy of Sciences’ Panel on Measuring Subjective Well-Being in a Policy-Related Framework. A.B.K. and A.A.S. are members of the OECD High-Level Expert Group on MultiLevel Subjective Well-Being. Financial support from the National Institute on Aging Royal grant P30AG024928 and grants R01AG042401, R01AG040629, and P01AG05842 are gratefully acknowledged.

10.1126/science.1256392