



International Society for
Quality-of-Life Studies

Working Paper No. 14

DOI:

<https://doi.org/10.5281/zenodo.17905297>

**Bridging the gap between
critical and applied
approaches in the science of
well-being - an analysis of
construct validity of
evaluative subjective well-
being measures**

December 2025

Luciano Sewaybricker

Title: Bridging the gap between critical and applied approaches in the science of well-being - an analysis of construct validity of evaluative subjective well-being measures

Luciano Sewaybricker (Postdoctoral researcher, University of São Paulo, Brazil)

Abstract: Assessment instruments, including those used to measure happiness and well-being, inevitably involve a degree of imprecision. Whether such imprecision is acceptable depends on the perceived utility of the results. This article examines the fragilities of evaluative subjective well-being measures—typically based on one or a few general questions—as a counterpoint to three common defences of their widespread use. First, we challenge the claim that non-definition of the object (a “non-paternalistic” approach) is desirable. While allowing respondents to interpret questions freely may enhance applicability, we argue that researchers still require minimal theoretical grounding and construct validity to justify what is being measured. This is especially critical when dealing with complex and volatile ideas like happiness or well-being, both in academia and among laypeople. Second, we critique the view that the absence of a clear definition is unproblematic as long as other psychometric properties (e.g., reliability or validity other than construct validity) are preserved. We show that such a stance underestimates the linguistic ambiguity and social embeddedness of terms like “happiness” and “life satisfaction,” which are governed by normative rules even when their meanings are unclear or unstable. Third, we examine the argument that, despite their conceptual weaknesses, evaluative subjective well-being measures remain useful. While we acknowledge their predictive power (e.g., job change, voting patterns), we emphasise that this often shifts attention away from happiness or well-being toward external outcomes. In normative applications—those aiming to raise happiness or well-being scores—the lack of construct clarity becomes especially problematic. From these critiques, we propose two steps to bridge the gap between theoretical-conceptual and more practice-oriented research: (1) making explicit the unavoidable adopted perspectives and their fragilities, and (2) paying greater attention to the use of words and expressions. Together, these recommendations aim to reconcile critical and applied approaches in the science of well-being, encouraging more theoretically robust and context-sensitive measurement practices.

Keywords: measurement, well being, evaluative subjective well being, life satisfaction, construct validity

Introduction

Assessment instruments—and the measurements derived from their use—inevitably tolerate a certain degree of imprecision. For instance, there is little discussion about discrepancies in time measurement across different clocks. This lack of concern is understandable: for the intended use of clocks, such variations are usually irrelevant—just a few seconds of difference per week or month among distinct instruments.

The same logic can be applied to assessments of happiness and well-being. Measurement imprecision must be weighed against the utility of the results for their intended purpose. Despite their imperfections, how useful are these measurements? This is the core question that should guide researchers’ efforts—broadly speaking—whether in revising theoretical frameworks, developing instruments, or applying existing ones in empirical studies.

Large-scale applications of these instruments are now common across psychology, economics, and related fields (O’Connor, 2023; Frijters et al., 2024; Helliwell et al., 2025). They argue that the field and its measurement tools have matured sufficiently to yield results that are highly valuable for both individual and collective decision-making, ranging from choosing activities and managing time to donating money, as well as evaluating public policies and government performance. As Adler and Seligman (2016, p.1) stated, “The science of well-being has evolved over the last 30 years, and well-being can now be reliably measured at both the individual and national levels.” Thanks to this advancement, we are said to be experiencing a “happiness revolution,” as Easterlin and O’Connor (2025) put it, likening the transformative potential of this research field to the industrial and demographic revolutions.

But the enthusiasm for these instruments has been accompanied by criticisms both of their theoretical consistency (see Alexandrova & Haybron, 2016; Alexandrova, 2017) and of the fragilities involved in translating such a

complex object into cardinal evaluation scales (particularly regarding heterogeneity in scale use and problems of linearity) (see Bond & Lang, 2019; Fabian et al., 2025).

This research aims to examine, in greater detail, the rigour of happiness and well-being assessment tools. We focus on evaluative subjective well-being (swb) measures —instruments that ask respondents to broadly assess their lives, such as the Cantril Self-Anchoring Striving Scale (Cantril Ladder) used in the Gallup World Poll and general life satisfaction questions, as found in the European Values Study (EVS). These tools are central to national and international monitoring initiatives and have been used to guide public policy (see UK HM Treasury, 2022).

Our argument is twofold. First, measurement imprecision alone does not invalidate an instrument. But examining present imprecisions reveals risks and highlights the need for theoretical clarity. Second, greater attention to theory can support the development of the field by clarifying the relation between evaluative swb measures and the broader concepts of happiness and well-being.

To advance this dialogue, we engage with three recurring arguments made by proponents of evaluative swb measures: (1) that leaving the object undefined is desirable; (2) that the lack of clarity about the object doesn't seem to be problematic; and (3) that these measures are useful despite their fragility. For each, we summarise the case, present our critique, and highlight limitations. Our aim is not to exhaust the debate but to reorganise it—bringing advocates of large-scale measurement into conversation with those calling for deeper conceptual work (see Flanagan et al., 2023; Alexandrova, 2017; Fabian et al., 2024; Alexandrova & Haybron, 2016).

Finally, a note on terminology: we use “happiness” and “well-being” in a deliberately broad sense. The definitions of each, as well as their relationship, are complex and have led to substantial divergence within the science of well-being (Lomas et al., 2024; Sewaybricker & Massola, 2022). This definitional plurality will be revisited later in the paper.

First argument: the non-definition of the object is desirable

The first argument we will explore is commonly referred to as the “non-paternalism” of evaluative swb measures. This argument highlights the desirability of leaving the object of evaluation deliberately broad. As O'Connor (2023, p. 3) explains, “In this way, respondents are able to evaluate what is relevant in their life, implicitly assign importance to each aspect, and convey one summary value.” In cases where the subject of inquiry is widely debated and lacks consensus, this non-paternalistic approach becomes particularly appealing (Sumner, 1996). To prevent the ultimate object of evaluation from reflecting what happiness or well-being means to the researcher—who may have developed a particular theory or instrument—rather than to the respondent, the evaluation parameters are intentionally left vague. In this sense, the instrument's response is taken as more meaningful in relation to the respondent's own life. As Diener et al. (1998, p.35) made the case:

With the increasing democratization and equality in the world, it is no longer defensible to rely only on kings and mandarins to prescribe the desirable life.(...) Thus, we allow people to decide whether their lives are satisfying based on their individual values, goals, and life circumstances.

This strategy of non-definition also supports the broad applicability of a given tool. Assuming that each of the many definitions of, say, happiness holds more relevance for some individuals than for others (see Joshanloo, 2025), a researcher who commits to only one such definition introduces significant obstacles to the application of their method. First, they will need to continuously justify why they have selected a particular definition of happiness over others (McMahon, 2006; Linton et al., 2016). Second, they must acknowledge that the broader the instrument's application (particularly when applied across cultural boundaries), the less likely it is that the chosen definition of happiness or well-being will resonate with respondents (Sewaybricker & Massola, 2025; Sollis et al., 2022).

To avoid these complexities, evaluative swb measures tend to employ broad phrasing, relying on expressions that are commonly used and assumed to be universal. For example, the Cantril Ladder asks respondents to indicate “where do you see yourself” in relation to the “best and worst possible life,” while the European Values Study (EVS) asks “how satisfied are you with your life as a whole.” These are expressions that refer to seemingly ordinary ideas, presumably familiar to nearly all participants. Terms like “satisfied with life” or a comparison with “the best possible life” are rarely treated as concepts in need of further definition—very different from more historically loaded terms like “happiness” or “well-being.”

In addition, the non-paternalistic strategy also relies on open-ended instructions, allowing participants to weigh whatever they find most relevant—be it, for example, cognitive, emotional, individual, collective, historical, or present-day aspects (Mahoney, 2023, p. 12). This is especially evident in the phrasing used in the EVS question, which explicitly instructs participants to respond “all things considered”.

As we can see, the strategy of not defining the object of evaluation greatly enhances the usefulness of these instruments: they are easier for participants to understand, avoid endless theoretical justifications and cultural divergences, and are therefore more practical for large-scale applications.

Lack of theoretical grounding and construct validity

In any case, even if the object of evaluation (which is supposed to be latent) is not explicitly stated in the wording of the instrument, it must be assumed that some understanding of it—however vague or tacit—underpins the instrument’s design and the entire investigative strategy (Tal, 2020). Without a construct, it would be impossible to justify the relevance of an instrument (“Why is it important to measure X?”) or the rationale behind its specific structure (“Why were certain formats, phrasings, or attitudinal items chosen?”). Thus, even if respondents are unaware that they are evaluating, for instance, their subjective well-being—since this is not named in either the items or the instrument title—the scientific discussion stemming from the results should revolve around this construct.

However, in the literature on evaluative swb measures, aside from brief definitions or nominal references, it is not easy to find well-developed articulations of the construct being assessed. One example is Loewenstein’s (2009) assertion that subjective well-being is “that which makes life worthwhile”. Other statements defend that “happiness is the ultimate goal in life and is self-evident” (Griffiths & Reeves, 2009) or that “Subjective well-being is thinking and feeling that your life is going well, not badly” (Diener, 2020). To some extent, non-paternalism appears to extend to the researchers themselves (see Alexandrova, 2018; Flanagan et al., 2023).

Still, this lack of conceptual clarity does not seem to pose a problem for many researchers, who indicate they are more focused on applying instruments and analysing the data produced, likely assuming that the construct is already sufficiently understood (Adler & Seligman, 2016; Diener, 2009). For example, in 1984, Ed Diener, even as he was compiling a fragmented research field, already emphasized the importance of focusing on measurement (Diener, 1984, p.543)—a perfectly reasonable stance, though somewhat ironic given the lack of agreement on the field’s core object (see Linton et al., 2016; Biswas-Diener et al., 2009; Margolis et al., 2021).

The emphasis on measurement, by itself, is not problematic. What might be problematic is the lack of clarity regarding the object of interest—especially if we cannot confidently claim that participants’ responses actually refer to the same object (i.e., the same construct). If that is the case, then the non-paternalism of evaluative swb measures becomes a serious issue. In the following section, we will start delving into the construct validity of these instruments, which broadly concerns assessing whether an instrument “measures what it purports to measure” (Nunnally & Bernstein, 1994, p.83).

Happiness or well-being as the ultimate goal

As previously noted, despite the abundance of data generated through the application of evaluative swb measures, their underlying constructs are often superficially or scarcely articulated. One of the few existing characterisations is that the object of interest is the ultimate goal of individuals:

We mean the thing that ultimately matters: your inner subjective state – the quality of your life as you experience it, how happy you are. We shall call this ‘well-being’ for short, but we always mean ‘subjective well-being’. Well-being, this book argues, is the overarching good and other goods (like health, family and friends, income and so on) are good because of how they contribute to our well-being. (Layard & De Neve, 2023, p.1)

This definition, which draws from intellectual traditions as diverse as Aristotelian ethics and utilitarian philosophy, calls for a depth of inquiry that is rarely pursued. To state that happiness or well-being is the ultimate end—that to which all other things aim or contribute (Layard, 2021)—easily becomes vacuous or tautological (Duncan, 2019, p.86). Beyond the assertion that happiness and well-being are “good” (by definition), little is actually known about the object itself. How can we determine whether a given sensation, circumstance, or thought qualifies as the ultimate goal? Without a more detailed articulation of what constitutes this ultimate end, the claim becomes circular: happiness is the ultimate end of human life; the ultimate end is called happiness. But how do we know what the ultimate end is, beyond simply naming it?

Utilitarian philosopher Jeremy Bentham (1823), for instance, argued that pleasure and pain are the “masters” of all people, and that happiness results from the net balance of the former over the latter. He also produced a long list of pleasures and pains, attempting to make their manifestations tangible (and thus measurable). Aristotle took a similar approach: the Greek philosopher detailed the expression of the ultimate goal as the practice of virtue and elaborated extensively on what virtues are and how they are cultivated to support his concept of eudaimonia (Aristotle, 1934). Both thinkers needed to present and commit to a particular nature of human beings before stating what the “ultimate goal” is. Simply saying that happiness or well-being is the ultimate goal contributes little to a meaningful understanding of the construct.

Happiness or well-being as being satisfied with life, or feeling that life is going well

In the case of presenting the object of interest in the form of "life satisfaction," "how well he likes the life he leads" (Veenhoven, 1984, p.22), or "a person's perception that life is going well" (Diener, 2020), the situation is somewhat different from the “ultimate goal” case. Due to the general familiarity with expressions like "satisfaction" or "doing well", this position might seem to avoid the need for theoretical clarifications. After all, we talk to one another using such terms, saying we are “satisfied with something,” that we’re “doing well,” or that life “is good.”

In this context, the reasoning presented by Veenhoven (1984, 2017) in favour of adopting this perspective is both illustrative of its influence on the field and the quality of the evidence it presents. Veenhoven's goal was to arrive at a workable concept amid the “discouraging confusion of tongues” regarding the object of interest, which he named happiness (1984, p.4). To this end, he recognised the need to make a motivated choice for a narrower concept; otherwise, his research “would probably get no further than repeating the old and unsolvable problems” of the happiness debate (Veenhoven, 1984, p.6–18). He concluded that among various understandings of happiness, the most operationalizable one was “the degree to which an individual judges the overall quality of his life-as-a-whole favorably.” (1984, p.22)

It is commendable that Veenhoven explicitly stated his interest in achieving an operationalizable definition of happiness and simultaneously acknowledged the cost (in his view, inevitable) of narrowing or simplifying the understanding of the object. This seems to be the aim of many researchers in the field. However, few offer such a

detailed and explicit presentation as Veenhoven: whatever the (theoretical) cost, a measurable object should be achieved.

Nevertheless, even if there is no doubt that this choice gives us a clearer operationalisation of the object, we are still left with three problems. First, whether the fundamental understanding of the object (e.g., Veenhoven's starting point and his framework) is adequate, given that there are divergences about what constitutes "happiness." Second, whether the result of such simplification (as it considers only one aspect of the object) can be regarded as a good representation of it. Third, whether instruments developed from the quantifiable aspect of the object are capable of eliciting its expression (the supposed latent content). Choosing a definition of happiness or well-being based on the criterion of "usefulness" may diverge from the criterion of "theoretical rigour" (see Basso & Alexandrova, 2025, for an example of that regarding inequality).

The fact that these expressions are used in everyday speech and that people report understanding the phrasing of such questions does not automatically imply the construct's validity. We will revisit this issue and its complexities later, together with psychometric support. For now, we can point out that, when framed in different ways, instruments appear to activate distinct mental processes, causing responses to correlate with varying factors. A classic example is the difference between Easterlin's (1974) findings and those of Stevenson and Wolfers (2008).

That said, although the wording of questions suggests different reflections and the correlations between results differ, these practical differences are not generally considered substantial. Helliwell & Wang (2012, p.15) illustrate this: "The bottom line of our comparisons among life evaluations is that when life satisfaction, happiness and ladder questions are asked about life as a whole, they tell very similar stories about the likely sources of a good life." This posture leads researchers to commonly group together data from different instruments to increase the amount of analysed information and reach stronger statistical results.

However, one problems arise from this posture. If this similarity of results can be read as consistency across instruments, it may also be interpreted as a sign of fragility in terms of construct validity. If, when asked about supposedly different objects (well-being, happiness, or life satisfaction) and in different formats (satisfied-unsatisfied or better-worse), people give similar answers, then perhaps the latent content under investigation is not actually being expressed.

Paternalism in instrument development and common sense

As McClimans (2013) pointed out, construct validity "asks how well a measure's outcomes compare to hypotheses derived from the theory that underpins the measure." Thus, the more opaque the object or reference theory is, the harder it becomes to discuss construct validity. In a field with a scarcity of theories or a preference for non-paternalistic approaches, it seems coherent that discussions about instruments tend to revolve around other elements, such as usefulness, common sense and other psychometric qualities.

The idea of commonsense is important for evaluative swb measures (Haybron, 2024), especially in light of the strong desire to evaluate an object with democratic contours (as opposed to an object defined and understood only by "specialists"). There are many aspects of human life where there is broad consensus on their being good or bad, practically requiring no ontological explanation for their value (or an underlying theory). For example, it's hard to imagine a scenario in which we would need to argue why being healthy is essential in someone's life. Consequently, health is widely regarded as essential for happiness or well-being. The same could be said for many other things: having a place to live, a stable income, freedom from food insecurity, or having people one can rely on. These consensuses are used by researchers both to present characteristics of the object and to assess the adequacy of instruments. Easterlin and O'Connor (2025, p.5) point this out:

A second way to assess the meaningfulness of happiness responses is by considering what people report when asked what makes them happy. (...) It turns out that the happiness of most people everywhere – in rich and poor countries, democratic and authoritarian – depends mainly

on the same three broad personal concerns – first of all, their economic situation, and then, family life and health.

Nonetheless, it's important to recognise that common sense plays a key role in the development of evaluation instruments by pointing toward relevant elements of the object of interest. Researchers, when choosing their object (e.g., happiness or well-being), must define which observable variables are assumed to correlate with the construct. Understanding common sense is crucial at this stage to evaluate whether, in the instrument's early uses, the construct behaves as expected (Alexandrova, 2017). For example, if a person's health worsens, their happiness or well-being is expected to decline (and if it doesn't, the instrument may need to be revised). The construction process of the Satisfaction with Life Scale (SWLS) by Diener et al. (1999, p. 277–278) illustrates this well. The instrument was validated based on its correlation with (among other things) absence of depression and fewer suicide attempts—two elements with little controversy regarding their relationship to happiness or well-being.

Still, the issue is not only whether a particular common-sense notion is relevant to happiness or well-being. It also concerns (a) recognising that not all common-sense notions are necessary for happiness and well-being, and (b) recognising that common-sense notions are generic and must be translated into the specifics of someone's life during the reflective process of evaluating happiness or well-being (Alexandrova, 2017). For example, “being healthy” can be expressed or perceived in very different ways: having the physical capacity (autonomy) to maintain one's hygiene, not suffering from any illness, or being able to run a half-marathon; the idea of being healthy for a child might be very different from for an old person. Without this level of detail, the common sense that health is an important element of happiness or well-being becomes impractical, despite the consensus.

Even with these points, it's fully understandable that common sense is considered during the initial development of instruments that aim to measure a complex construct. Common sense and measurement attempts can help better understand the construct in an iterative process (Rodriguez Duque et al., 2024). Nevertheless, it is desirable for researchers to be aware of the fragile relation between common sense and theory/construct. In the case of SWLS validation, for instance, it would be important for common sense indicators (like absence of depression or fewer suicide attempts) to be explicitly derived from a richer theoretical framework about what happiness or well-being consists of, or to serve to enrich that framework. Otherwise, we would merely know that suicide incidence should be inversely related to SWLS scores, but we wouldn't understand why that should be the case. The absence of such explanations severely limits discussions of construct validity.

Ahmed (2023) also points to the risk of treating common sense uncritically, particularly when it is left unarticulated. To do so may conceal the value judgements at play, making it more difficult to examine their origins and effects. Why are these forms of common sense, rather than others, used to assess the adequacy of the instrument? What assumptions about common sense were deemed appropriate? Is it common sense for whom? Thus, defending the non-paternalism of evaluative well-being instruments based on common sense does not indicate an absence of value judgments—it only avoids engaging in discussion about them (see Rodriguez Duque et al., 2024, p. 3).

Second argument: the lack of a clear definition is not a problem

The second argument can be understood as a response to the challenges related to construct validity mentioned before. Broadly speaking, this argument holds that the absence of a precise definition of the object-construct is not necessarily a problem. If people generally report that they are evaluating “happiness” or “well-being,” then researchers need not be entirely sure about the object's boundaries. In this line of reasoning, other psychometric properties of the instruments are considered strong enough, as illustrated by Diener et al. (1999):

These measures do possess adequate psychometric properties, exhibiting good internal consistency, moderate stability, and appropriate sensitivity to changing life circumstances. Furthermore, global reports show a moderate level of convergence with daily mood reports,

informant reports, spouse reports, and recall for positive versus negative life events. People who score high on global life satisfaction are less likely to attempt suicide and to become depressed in the future. (Diener et al., 1999, p. 277–278)

Let us break this argument into three supporting elements. The first concerns the importance of large samples (Easterlin & O'Connor, 2025). That is, even if there are variations in what respondents consider when answering evaluative well-being instruments (e.g., different notions of happiness or well-being), the large sample size would mitigate the impact of these variations and yield a reliable average. This same logic underpins much of cross-cultural research in the science of well-being, justifying comparisons between vastly different countries—as is the case with the annual ranking published in the World Happiness Report (Helliwell et al., 2025). The underlying idea is that we might not know specifically the meaning of the construct for one particular person, but we have a good enough understanding for people in general.

The second support derives from people commonly making everyday life evaluations similar to those required by evaluative swb measures. People ask themselves and answer each other's questions as to whether life is going well, whether they are satisfied with something, or even whether they are happy (Kahneman & Riis, 2005). If this is a common occurrence, it would suggest that people are sufficiently familiar with the object to assess their relationship with it in an instrument.

Finally, the third supporting element refers, as highlighted in the earlier quote from Diener et al. (1999), to the psychometric properties of instruments. High response rates (Veenhoven, 2017), strong test-retest scores, and the alignment between self-assessments and the assessments of others who know the respondent (such as partners and friends; Schneider & Schimmack, 2009) are commonly cited as evidence that respondents know what they are evaluating (Frijters et al., 2024, p. 4).

Assuming participants understand the object is a big leap

All three of these supports are consistent in their own right and widely endorsed within the field. However, even if we temporarily set aside concerns about construct validity, we argue that these supports should be taken with a grain of salt. There are alternative explanations that deserve closer attention.

Let us begin with the first supporting element, concerning the importance of large samples. Here, we offer two counterpoints: “ignoring deviant cases” and “simplification without meaning.” Panasiuk et al. (2025) and Barrington-Leigh (2025) offer interesting examples of the former. By analysing databases from different studies that used single-question well-being assessments, both studies found that the distribution of responses differs significantly from that of a normal curve. More specifically, Barrington-Leigh (2025) pointed out that the difference in response distribution is more pronounced in some regions than in others. In European and Anglophone countries, where most theories and measures are developed, the variation in the distribution of responses between different instruments is considerably smaller than in other regions. This heterogeneity or homogeneity in response patterns may be explained by differences in familiarity with the semantic field related to happiness and well-being (which we will address later), in familiarity with numerical evaluations of subjective experiences (such as school grading or hospital pain scales), or even in familiarity with quantifying what is considered important (a practice that is uncommon for many Indigenous peoples, for example—see Sewaybricker & Massola, 2025; Da Silva Sinha et al., 2017).

These studies reinforce a point that is often overlooked yet obvious. Discussing results based on averages may create a false impression of consensus and that common sense is being reinforced. For certain purposes, such as some public policy evaluations, this simplification (ignoring deviations from the average) may be justifiable due to its practical utility. However, averages should not be taken as support for an instrument's rigour. In particular, paying great attention to response heterogeneity can be especially relevant for deepening our understanding of the

construct and advancing theory. This will be especially important the more diverse the groups being compared (such as a comparison for a whole country or between countries).

The second counterpoint, "simplification without meaning," concerns the potential consequence of evaluative swb measures tending to treat their object as simple. The simpler the object of interest is presumed to be, the easier it becomes to develop a "rigorous" instrument for its evaluation. As previously mentioned, this logic was utilised (though not explicitly framed this way) by Veenhoven (1984, p.4) when he reduced happiness to "the overall appreciation of one's life-as-a-whole." This is almost a necessity if researchers intend to apply an instrument to a large population.

For this reason, it is not surprising that the large samples obtained by evaluative swb measures result from a qualitatively simple process when compared, for example, to interviews or questionnaires with open-ended questions. Everything participants might wish to express about the construct is narrowed down to the provided options (selecting one among a few numerical or verbal alternatives that express gradations) in a reduced amount of time. It is assumed that different people selecting similar responses are expressing the same or similar things. The "5" a person chooses in response to a given stimulus (item or question) is considered comparable to the "5" selected by another (Fabian et al., 2025).

This simplification found in evaluative swb measures is not accidental; it follows a long process of testing and evaluation of psychometric properties (Diener et al., 1985; Alexandrova & Haybron, 2016; Alexandrova, 2017). Nonetheless, it is crucial to emphasise the high degree of simplification employed when addressing a historically complex subject (McMahon, 2006). Both the EVS and the Cantril Ladder use a single attitudinal item, usually answered in under ten seconds (Liu & Netzer, 2023; Samuelsson et al., 2024). Evaluative swb measures appear to elicit a shortened (heuristic) form of reflection that is seldom discussed and may have little—or only superficial—relation to the object of interest, even if they display good psychometric properties.

More importantly, debating or even noticing oversimplification can be challenging, as the nature of the construct itself is inherently nebulous. Are people truly evaluating happiness or well-being, or merely a simplistic version of the construct? Or perhaps we are dealing with a different—new—construct altogether? It is challenging to begin answering these questions without greater attention to theoretical foundations.

Linguistic complexity

The second support concerns the regular use of the words and expressions present in evaluative swb measures in everyday language. The standard argument is that this frequent usage suggests that people are familiar with both the object being addressed and the reflective process required by these instruments.

Regarding familiarity with the object, Cromby (2011) offers an alternative explanation grounded in linguistics. Cromby argues that regular use of a word should not be interpreted as evidence that individuals possess sufficient and stable knowledge of the object to which the word refers. Instead, regular use of a word primarily indicates that the person knows the social rules governing its use, independent of any stable or deep understanding of the object itself. For example, when we wish someone a "happy birthday," we do not need to rely on a specific definition of happiness. Nor does the person receiving the wish—typically responding with a "thank you"—need to have a formed idea of what happiness is or to understand the particular perspective of the one offering the wish. We know that "happiness" is something positive, desirable, and appropriate to mention to someone on their birthday—and that suffices.

Understanding the complexity of language—its comprehension and expression—casts doubt on the assumption that individuals are either familiar with the reflective process required by questionnaires or know the object of inquiry well. An alternative hypothesis is that people are familiar not with evaluating their lives in terms of the construct (happiness or well-being), but rather with solving the social problem of appropriately using expressions

like "happiness." People's reflective process may be more guided by thoughts like: "Given the social context I am in, would it be appropriate to say I am happy or not?"; "What would be the most socially acceptable answer?"; "What are the consequences of saying I am or am not happy?"; "How will others react to my answer?"

There is yet another argument against the idea that people are familiar with happiness or well-being and their evaluation. This argument, rather than linguistic, is logical in nature and centres on a common issue in evaluative swb measures: is it truly possible for someone to rigorously and stably evaluate life as a whole? There is extensive discussion on this matter, well-known among proponents of evaluative swb measures. Kahneman (2005), for instance, investigated the effects of memory on self-assessment, illustrating how volatile retrospective evaluations of life can be. In the philosophical field, White (2006) argued that treating "life as a whole" as a unit of analysis is logically impossible. It would be impossible to consider everything pertaining to life simultaneously, as new elements and perspectives will always emerge.

Psychometric properties with a blurred construct

In light of the weaknesses indicated earlier, let us revisit the third support element, which states that an imprecisely defined object by researchers would not be problematic: the psychometric properties of the instruments would indicate that participants know what they are evaluating. The numbers regarding these psychometric qualities are indeed significant and represent potentially strong elements in defending an instrument. However, once again, there are alternative hypotheses that explain why such numbers are so significant—hypotheses that do not involve the respondents having (explicit or tacit, conscious or unconscious) knowledge about the object of interest.

One commonly presented point is the high response rate. According to Veenhoven (2017), fewer than 1% of participants fail to respond to the questions or mark alternatives like "don't know." In other words, it is assumed that if the vast majority of people complete the tests adequately, it is because they possess explicit or tacit knowledge of the object of interest. However, as mentioned earlier, for such a highly complex object, which is hard for even specialists to grasp, the high response rate may reflect people's familiarity with its expressions rather than with the concept (Cromby, 2011). The diversity of definitions about happiness and well-being (see McMahon, 2006; Haybron, 2007; Sewaybricker & Massola, 2025), of instruments (see Linton et al., 2016; Veenhoven, 2025), and the general difficulty people have in presenting a coherent, stable, and well-defined idea of this object seem to support this hypothesis (Fabian et al., 2025).

In any case, regardless of whether people have formal and explicit knowledge of the object of interest, a defender of evaluative swb measures might argue that the high response rate stems from tacit knowledge about the object. People filling out the instruments would know what the object is, even if they cannot explain it or justify the score. This could also help explain why response time for evaluative swb measures is usually less than 10 seconds. However, this is a weak argument for supporting the rigour of an instrument, as we would have no way to verify it. If researchers do not clearly know what the object is and participants can't explain it, how can we ensure that the knowledge elicited by the instrument is about happiness or well-being?

We can extend this same problem to the test-retest score and to the alignment between self-assessments and assessments by others who know the respondent. Given the high complexity of the object of evaluation (happiness or well-being), and perhaps even the impossibility of evaluating life as a whole, it would not be unreasonable to suppose that the assessment is based less on this latent content and more on common sense and general rules governing the use of expressions. Both common sense and general rules of use are more stable and observable, thereby increasing cohesion both among individuals and within the same individual over time. We are more familiar with attributing the adjective "happy" to people and situations than to the definition of happiness. Thus, we might not know what happiness or well-being is, but we have a reasonable understanding of who is generally considered to be happy (e.g., those who smile more often or appear to be in good health) and of the situations or events that are typically associated with happiness (e.g., earning a job promotion, getting married, or having children).

Third argument: despite their fragilities, evaluative swb measures are useful

The third and final argument in defence of evaluative swb measures that we will address concerns the usefulness of these instruments. That is, the weaknesses pointed out so far will only be relevant—or worth deepening—insofar as they affect the usefulness of these assessments. This is a remarkably consistent line of reasoning, as it sidesteps the weaknesses derived from definitions of happiness or well-being (construct validity).

The usefulness of the evaluative swb measures will be thought mainly as the predictive power and the strength of correlations with variables of interest. The underlying thinking is roughly as follows: if instruments can predict behaviours and show strong correlations with relevant variables, then there is no need to delve into the details of what happiness or well-being actually are—or even to question whether the instruments genuinely measure happiness or well-being.

By focusing on usefulness, research efforts can be viewed as twofold. First, the treatment of data derived from different instruments to make their results comparable (e.g. Kaiser & Oswald, 2022; Kaiser & Lapinteur, 2025; Frijters et al., 2024). This is a necessary effort to work with large datasets in a field populated by instruments with diverse statements and scales. The second effort involves investigating the respondents' reporting function, a necessary step for adjusting instruments and research design to promote greater comparability both across participants and within the same participant over time (e.g., Prati & Senik, 2025).

These efforts are not free from disagreement. Divergences emerge especially in the discussion on the reporting function, since it involves a complex “affective, cognitive, social, and linguistic process that maps subjective judgements to numerical response categories on a scale question” (Fabian et al., 2025).

In this aspect, some of the most incisive critiques of evaluative swb measures have been presented by Bond and Lang (2019), Fabian (2022), and Fabian et al. (2025). These works emphasise the variability of the reporting function. In particular, they highlight weaknesses arising from the very nature of the object of study itself (Flanagan et al., 2023; Alexandrova, 2017; Angner, 2013; Rodriguez Duque, 2024). Unlike a ruler, which measures length using consistent units, happiness or well-being—an abstract and highly contested object—is measured using integers, a property of an entirely different kind. A score of 5 may mean a great deal—or very little—to different people, or even to the same individual at different times. Thus, score increases may not correspond to linear increments of the underlying concept. For instance, the jump from 5 to 6 may feel larger than from 8 to 9. This problem is not unique to the cardinal evaluation of happiness or well-being, but it is more acute in this case, given the complexity of the object.

It is important to emphasise that these critiques are motivated by an interest in happiness and well-being and are grounded in a particular understanding of the object—from which the fragility of the relationship between construct and measurement is derived. By contrast, for proponents of the third line of argument, the weaknesses of evaluative swb measures are primarily interpreted in terms of the usefulness of the instrument (its predictive power and correlations), rather than in terms of the complexity of the object itself. Research following this orientation tends to focus on maximising the usefulness of instruments, engaging in debates about how to reduce variation while avoiding discussion of the central object. For example, it is common to assume equivalence between the construct (happiness or well-being) and the responses elicited by evaluative swb measures (see Easterlin, 1974). In other cases, exact equivalence is not assumed, but not because the object of interest would not be expressed; rather, the variations in the reporting function are seen as introducing noise into its expression (see Prati & Senik, 2025).

As a result, identifying and reducing noise becomes a central endeavour. Strategies have been developed to minimise heterogeneity in the reporting function, such as identifying participants' response styles and adjusting their answers accordingly (Kaiser, 2022). Other methods include adding diagnostic questions or using anchoring vignettes (King et al., 2004; Weiss & Roberts, 2018). Prati and Senik (2025) propose another calibration strategy to improve temporal comparability within the same respondent. Inspired by the practice of adjusting financial data

for inflation, they suggest including retrospective evaluations (e.g., “How satisfied are you with your life today?” and “How satisfied were you with your life at the time of the previous survey?”). In this way, the most recent score can be adjusted according to the frame of reference of the earlier scale. The advantage of this correction is that it also addresses the “ceiling effect” and the potential limit to how much happiness or well-being a person can experience (see Schutte et al., 2019; Fabian et al., 2025). Since the current score is recalibrated relative to the previous evaluation, it becomes possible to conceptualise unlimited increments in scale values.

Regarding linearity, there is evidence that the impact of the fragilities pointed out by some of the critics of evaluative swb measures is far less significant than expected. People seem to use happiness or well-being scales in a cardinal manner (Layard & De Neve, 2023; Prati & Senik, 2022), especially in scales with more than three response items (Kaiser & Vendrik, 2020). As Kaiser and Lapinteur (2025, p.1) wrote, “respondents use survey scales in ways that deviate from linearity, but only mildly so.”

In conclusion, human beings can apparently operationalize a reliable numerical scale for unmeasurable inner emotions. How they do this, using made-up integers on a scale that does not truly exist, is currently unknown. It demands further scientific investigation. (Kaiser & Oswald, 2022, p.5)

Nonetheless, it is essential to emphasise once again the contrast with the more critical stance. For critics, the weaknesses are perceived as inherent to the object’s characteristics. From this perspective, evidence of “quasi-linearity” might be interpreted not as a validation of evaluative swb measures but as an indication that they capture something other than happiness or well-being – as, by definition, these are non-cardinal and variable constructs (Alexandrova, 2017; Rodriguez Duque et al., 2024; Angner, 2013). The question of what latent content is actually being expressed in evaluative swb measures, therefore, remains open to debate.

Evaluative swb measures guiding action: agnostic and normative approaches

As mentioned, various studies indicate that, regardless of construct validity and other psychometric qualities, evaluative swb measures are good at predicting people’s behaviour (Kaiser & Oswald, 2022; Ward et al., 2021; Liberini et al., 2017). Kaiser and Oswald (2022), after examining current assessments in Germany, the UK, and Australia, found that these instruments can predict “exit behaviour” or “get-me-out-of-here” actions, such as changing jobs, moving to a new house, going to the hospital, or ending a marriage. Studies by Ward et al. (2021) and Liberini et al. (2017), which focused on the political context, found that evaluative swb results predict voting patterns.

Given their predictive power, it’s unsurprising that evaluative swb measures are used by individuals and organisations to guide decisions (for work-related discussions, see De Neve & Ward, 2025; for public policies, see UK HM Treasury, 2022). For instance, companies may implement measures to improve employee retention and reduce turnover; governments may develop programs to address a potential rise in unemployment; politicians may adjust their campaign strategies based on public satisfaction or well-being levels.

Let’s call “agnostic” the approach in which, without engaging in conceptual debates, one supports using evaluative swb measures for decision-making. In such examples, the results guide actions based on trends they reveal (predictive power). That is, the decision-maker need not understand happiness or well-being—or even accept researchers’ definitions. They may not even know what happiness is. It suffices that they trust the statistical analyses and see value in the predictive behaviours and correlations. Thus, it’s hard to find arguments to refute the agnostic approach’s claims, as it bypasses the fragilities of construct validity.

A caveat, however, is worth noting. In this stance, happiness or well-being ceases to be the research focus. For example, the focus shifts to unemployment, rather than happiness itself. Thus, the agnostic approach requires no conceptual deepening and need not engage in multidisciplinary or theoretical debate. On the other hand, studies interested in why these predicted behaviours occur—which involves understanding happiness—operate under a very different logic from agnostic ones.

To illustrate, when correcting for the “inflation” of evaluative swb measures and testing their strategy with income (GDP) in the United States, Prati and Senik (2025) do not raise questions about whether, why, or how income (GDP) should be correlated with happiness or well-being. They assume that such a correlation is to be expected and that it would serve as an indicator of the quality of their proposed intertemporal adjustment. In their conclusion, they write:

If it is true, as Kahneman and Krueger (2006, p.18) suggest, that “the invariance of reported satisfaction to material living standards and major changes in life circumstances is a reason why some economists are skeptical of the validity of measures of subjective well-being”, then our study helps to make progress on this front. (Prati & Senik, 2025, p34).

Cultural particularities that might bring certain groups closer to (or further from) the response patterns typically observed (e.g., linearity) are not of direct concern to agnostics—except insofar as they affect the predictive power or the strength of correlations with other variables of interest.

There is yet another perspective that seeks to influence actions based on evaluative swb measures. Let’s call this the “normative” approach. Unlike the agnostic approach—focused on predicted behaviour and correlation—the normative approach is interested in the actual scores of evaluative swb measures (the self-declared nominal level of the construct) (Frijters et al., 2024; Plant et al., 2025). Specifically, it is concerned with how to increase these scores, as can be seen in the focus of specific institutions (e.g., Action for Happiness and the Happier Lives Institute) and in the work of those who argue that going “beyond GDP” requires placing the promotion of well-being or happiness at the centre of policy (Veenhoven, 2017; Weijers & Jarden, 2013; Frey & Stutzer, 2017).

The term “normative” reflects the value judgment inherent in this position. Unlike the agnostic approach, which doesn’t need to rely on a specific idea of the object being measured, the normative approach is dependent on the value and meaningfulness of it: instruments measure something good (happiness or well-being), and higher scores are desirable. Without this assumption, there would be no reason to seek to raise scores.

Although the normative approach engages with the same debates on the reporting function as agnostics (e.g., in relation to Well-being-Adjusted Life Years), it cannot avoid grappling with the conceptual complexity of happiness and well-being. It must engage with debates about construct validity. The possibility—highlighted by critics—that the latent content expressed in evaluative swb measures may not in fact be happiness or well-being poses a serious challenge to the entire normative framework. Among other things, it requires reflecting on whether “corrections” to participants’ scores, such as those proposed by Prati and Senik (2025), might in fact distort the latent content. Consequently, when operated within a normative logic, adjustments to the reporting function for the sake of comparability and linearity create as many problems as they solve.

Discussion

Tools and theories are not perfect

Concepts and their respective assessment tools will never be perfect in the sense of definitively resolving all existing disagreements within a field of study. Whether in discussions about the concept itself and what it consists of, the best ways to represent or express it for evaluation, or the design of the assessment instrument, there will always be room for debate. There is no “holy grail” in the assessment of well-being or happiness—and there never will be. In fact, one could argue that, when it comes to mental states, these instruments are even further from seeming to be a “holy grail.” This differs, as previously mentioned, from the natural sciences, such as measuring time with clocks or length with a ruler. The measurement of an external, directly perceptible object facilitates consensus among researchers. That said, even though there are particular complexities in understanding happiness or well-being, this does not mean that we cannot strive for greater rigour and develop useful constructs and instruments at the same time.

But it is essential to acknowledge the limitations of research so that results are not extrapolated beyond what they rigorously indicate, thereby avoiding the risk of promoting injustice. Therefore, we believe that research using evaluative swb measures would benefit from greater dialogue between those who support their use and those who are critical of it. We understand that such dialogue will be fostered, in part, by greater clarity regarding the intentions and theoretical foundations of those involved in the field. Our aim in this article has been to contribute to this by offering an interpretation of the interests and intentions at play.

In addition, however, we also believe that dialogue will be facilitated by changes in how research is communicated. Based on the discussion above, we propose two suggestions to advance this. The first is to make explicit the unavoidable perspectives adopted and their limitations. The second is to pay greater attention to the language used—the words and expressions chosen. These suggestions aim to enhance clarity regarding the object of evaluation and, consequently, to strengthen the dialogue between theoretically oriented and more practice-oriented research.

Recognising and stating perspectives and fragilities

One obstacle to achieving rigour in research is the presentation of brief and vague definitions of the object of interest. This aspect hinders debates focused on concepts and theories. If it is unclear what the object of interest is and which theoretical perspectives are being adopted, how is it possible to construct a meaningful critique? In such cases, the dialogue is likely to be fruitful only when limited to methods and results.

Thus, while for those in favour of evaluative swb measures this is already common at the research design and data analysis level, it is equally important that it becomes a standard practice at the theoretical level. This involves acknowledging the normative choices of the research—arbitrary decisions that may be embedded in the adoption of specific theoretical frameworks, the selection of particular instruments, the framing of questions, or the suppositions about the object being considered for testing the quality of an instrument.

In other words, even if one wishes to sidestep the complexity of the object or allow participants to use their own criteria in evaluative swb measures, researchers will inevitably (explicitly or not) adopt certain assumptions about happiness or well-being. Each of these choices represents particular interests and values, introducing inevitable fragilities—ones that researchers should examine and, where possible, make explicit.

This recognition opens the door for theoretical–conceptual research to be presented as complementary rather than antagonistic to practice-oriented research. Two examples of potential complementarity are studies that qualitatively explore the diversity of the reporting function (see heuristics and cognitive interviews, Kahneman, 1992; Fabian et al., 2025) and studies that develop happiness or well-being assessment instruments tailored to localised or small-scale populations (see participatory-approach and democratic development of instruments, Sollis et al., 2022; Yap & Yu, 2016; Alexandrova & Fabian, 2022).

Rigour in the use of terms and expressions

Our second suggestion pertains to a more rigorous use of terms and expressions. We make this recommendation acknowledging the potential difficulty in implementing it, since the choice of words in research, articles, and reports involves not only precision but also the desire to be understood by an audience. Moreover, we are dealing with a particularly confusing semantic field in which word choice often requires consideration of translation, particularly into English. For example, due to the linguistic and conceptual complexities of the object (happiness or well-being), Easterlin and O'Connor (2025, p. 12) proposed a practical solution: “We refer to them both by the less cumbersome and more user-friendly term, ‘happiness.’” However, we might ask: for whom should it be less cumbersome and more user-friendly? While such a choice may be convenient, it may hinder dialogues grounded in conceptual rigour.

In this sense, it seems important to consider words or expressions in their original contexts, respecting—where possible—the word used in the local language of the research, either by retaining the original term or noting the translation. For instance, while the English word happiness underwent a process of trivialization over the 20th century (Oishi et al., 2013; Wierzbicka, 2004; Sewaybricker & Massola, 2022), this does not mean that words from other languages commonly translated as happiness have gone through the same process (as Wierzbicka argues in the case of the French “bonheur”). So, even when words are part of the same semantic field, they should not be treated as synonyms. This also applies to terms used by historical authors (e.g., Aristotle and *eudaimonia*, Epicurus and *makarios*, Jeremy Bentham and happiness). Moreover, investigating the etymology and historical transformation of words can help ensure more precise usage. Such inquiries can help avoid inadvertently replicating arguments like that of Ed Diener, who introduced subjective well-being instead of happiness to gain credibility in the scientific community, not due to any specific theoretical or historical justification (Disabato et al., 2025, p. 2; Lyubomirsky, 2008, p. 316).

Given that the object is complex to define, polysemic, and contested in the literature, it is still possible to strive for internal consistency within each research study. A straightforward way to ensure internal rigour is to state clearly how expressions will be used (e.g., whether happiness and well-being will be treated as synonyms). But it is valuable to go further. It is common for papers to cite many other studies that don’t treat the object of interest the same way or use the same terminology. Thus, it is essential to double-check how expressions are used and justify the chosen conceptual approximations. For instance, it would be careless to treat happiness and well-being as synonyms without acknowledging that cited studies also do the same.

Conclusion

The growing relevance of evaluative swb measures and the expanding body of research employing them reflect the value being attributed to the assessment of happiness or well-being. However, this is an object of particular complexity, which has led many researchers to define it in terms of what can be measured (e.g., Veenhoven, 2000). In this case, the instrument ultimately determines the object. If happiness or well-being exceeds what is being measured, this is not considered a problem for many researchers.

This led Alexandrova and Haybron (2016, p.1103, emphasis in original) to write that construct validation in well-being science is “*theory avoidant*, sacrificing valid theoretical knowledge for statistics for no good reason.” This posture hinders broad discussions of scientific rigour by excluding the complexity of the object from the “discussion table” — which might make sense for those adopting an agnostic stance, but not for others primarily interested in happiness or well-being.

This practice recalls the anecdote of someone who searches for a lost item only where the light is shining (Angner, 2011). Addressing the same context, Ahmed (2007, p.9) warned that advancing without clarity about the object may lead researchers to confirm their biases and find happiness where they expect it to be. The fact that the conclusion of Easterlin’s (1974) research is called a “paradox,” for instance, already indicates an assumption that increases in happiness and income should progress together.

It is essential for the development of the field—and for the development of instruments themselves—that we give greater importance to the level of constructs and theory; that we bridge the gap between measurement efforts and critical discussion. If we fail to do so, we risk continually simplifying the object of interest, reducing it to the most tangible and measurable elements, even if these are not the most meaningful, rigorous, or ethically responsible. The current moment, referred to as a “happiness revolution” by Easterlin and O’Connor (2025), could either become a revolution aligned with scientific rigor, or merely a discursive one—a likely risk, according to Flanagan et al. (2023, p.34): “We predict that the idea that government should promote happiness will shortly be seen for what it is: a fad, a passing fancy, a branding overreach.”

References

- Adler, A., & Seligman, M. (2016). Using well-being for public policy: Theory, measurement, and recommendations. *International Journal of Wellbeing*, 6(1), 1–35. <https://doi.org/10.5502/ijw.v6i1.429>
- Ahmed, S. (2007). The happiness turn. *New Formations*, 63, 7–14. Retrieved November 19, 2024, from <https://journals.lwbooks.co.uk/newformations/vol-2007-issue-63/>
- Ahmed, S. (2023). *The feminist killjoy handbook: The radical potential of getting in the way*. Hachette UK.
- Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford, UK: Oxford University Press.
- Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford University Press.
- Alexandrova, A. (2018). Can the science of well-being be objective? *The British Journal for the Philosophy of Science*, 69(2), 421–445. <https://doi.org/10.1093/bjps/axw027>
- Alexandrova, A., & Haybron, D. M. (2016). Is construct validation valid?. *Philosophy of Science*, 83(5), 1098–1109. <https://doi.org/10.1086/687941>
- Alexandrova, A., & Fabian, M. (2022). Democratising measurement: Or why thick concepts call for coproduction. *European Journal for the Philosophy of Science*, 12(7), 1–23.
- Angner, E. (2011). Are subjective measures of well-being ‘direct’?. *Australasian Journal of Philosophy*, 89(1), 115–130. <https://doi.org/10.1080/00048400903401665>
- Angner, E. (2013). Is it possible to measure happiness? The argument from measurability. *European Journal for Philosophy of Science*, 3, 221–240. <http://dx.doi.org/10.1007/s13194-013-0065-2>
- Aristotle (1934). *Aristotle in 23 Volumes* (transl. H. Rackham). Cambridge, MA: Harvard University Press.
- Barrington-Leigh, C. (2024). The econometrics of happiness: Are we underestimating the returns to education and income? *Journal of Public Economics*, 230(1): e105052. <https://doi.org/10.1016/j.jpubeco.2023.105052>
- Barrington-Leigh, C. (2025). Are international happiness rankings reliable?. *arXiv preprint* <https://arxiv.org/html/2509.06867v1>
- Basso, A., & Alexandrova, A. (2025). Measurement requires compromises: the case of economic inequality. *Studies in History and Philosophy of Science*, 113, 88–97. <https://doi.org/10.1016/j.shpsa.2025.08.005>
- Bentham, J. (1823). *An Introduction to the Principles of Morals and Legislation*. Oxford, UK: Clarendon Press.
- Biswas-Diener, R., Kashdan, T. B., & King, L. A. (2009). Two traditions of happiness research, not two distinct types of happiness. *The Journal of Positive Psychology*, 4(3), 208–211. <https://doi.org/10.1080/17439760902844400>
- Bond, T. N., & Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, 127(4), 1629–1640.
- Cromby, J. (2011). The greatest gift? Happiness, governance and psychology. *Social and Personality Psychology Compass*, 5(11), 840–852. <https://doi.org/10.1111/j.1751-9004.2011.00398.x>
- Da Silva Sinha, V., Sinha, C., & de Araujo Sampaio, W. B. (2017). The Many Ways to Count the World: Counting Terms in Indigenous Languages and Cultures of Rondônia, Brazil. *Brief Encounters*, 1(1). <https://briefencounters-journal.co.uk/article/id/177/>
- De Neve, J. E., & Ward, G. (2025). *Why Workplace Wellbeing Matters: The Science Behind Employee Happiness and Organizational Performance*. Harvard Business Press.
- Diener, E. (1984) Subjective Well-being. *Psychological Bulletin*, 95(3), 542–75. <https://doi.org/10.1037/0033-2909.95.3.542>
- Diener, E. (2020). Happiness: the science of subjective well-being. In R. Biswas-Diener & E. Diener (Eds), *Noba textbook series: Psychology*. Champaign, IL: DEF publishers. Retrieved from <http://noba.to/qnw7g32t>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The life satisfaction scale. *Journal of Personality Assessment*, 49(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
- Diener, E., Sapyta, J. J., & Suh, E. (1998). Subjective well-being is essential to well-being. *Psychological Inquiry*, 9(1), 33–37. https://doi.org/10.1207/s15327965pli0901_3
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125(2), 276–302. <https://doi.org/10.1037/0033-2909.125.2.276>
- Disabato, D. J., Goodman, F. R., & Kashdan, T. B. (2025). The hierarchical framework of wellbeing (HiFWB). *Frontiers in Psychology*, 16, 1515423. <https://doi.org/10.3389/fpsyg.2025.1515423>
- Duncan, G. (2019). Happiness and the new politicization of subjectivity. In N. Hill, S. Brinkmann, & A. Petersen. *Critical happiness studies* (pp. 83–97). Routledge.

- Easterlin, R. A. (1974). Does Economic Growth Improve the Human Lot? Some Empirical Evidence. In: David, P. A. & Reeder, M.W. (eds.). *Nations and Households in Economic Growth*. Academic Press, New York, pp. 89-125.
- Easterlin, R. A., & O'Connor, K. J. (2025). *The Happiness Revolution in Europe*. Cambridge University Press.
- Fabian, M. (2022). Scale norming undermines the use of life satisfaction data for welfare analysis. *Journal of Happiness Studies*, 23(4), 1509–1541. <https://doi.org/10.1007/s10902-021-00460-8>
- Fabian, M., Alexandrova, A., Coyle, D., Agarwala, M., & Felici, M. (2024). *Wellbeing public policy needs more theory*. <https://doi.org/10.17863/CAM.112205>
- Fabian, M., Kaiser, C., Panasiuk, S., Funk, S., Pountney, L., Brett, C. (2025). Evidence against the simple validity of life satisfaction scales from long cognitive interviews. forthcoming.
- Flanagan, O., LeDoux, J. E., Bingle, B., Haybron, D. M., Mesquita, B., Moody-Adams, M., ... & Wilson, Y. Y. (2023). *Against happiness*. Columbia University Press.
- Frey, B. S., & Stutzer, A. (2017). *Public choice and happiness (No. 2017-03)*. CREMA Working Paper. Retrieved November 19, 2024, from <https://hdl.handle.net/10419/214596>
- Frijters, P., Krekkel, C., Sanchis, R., & Santini, Z. I. (2024). The WELLBY: a new measure of social value and progress. *Humanities and Social Sciences Communications*, 11(1), 1-12. <https://doi.org/10.1057/s41599-024-03229-5>
- Griffiths, S., & Reeves, R. (2009). *Well-being: How to lead the good life and what government should do to help*. The Social Market Foundation.
- Haybron, D. M. (2007). Do we know how happy we are? On some limits of affective introspection and recall. *Nous*, 41(3), 394-428. <https://doi.org/10.1111/j.1468-0068.2007.00653.x>
- Haybron, D. M. (2024). Well-being policy: Consensus hallmarks and cultural variation. In J. D Rooney & P. Zoll. *Beyond Classical Liberalism* (pp. 205-221). Routledge India.
- Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2025). *World Happiness Report 2025*. University of Oxford: Wellbeing Research Centre.
- Helliwell J. F., & Wang S. (2012). The state of world happiness. In Helliwell, J. F., Layard, R. & Sachs, J. (eds.) *World Happiness Report 2012*. New York: UN Sustainable Development Solutions Network.
- UK HM Treasury. (2022). *The Green Book – Central Government Guidance on Appraisal and Evaluation*. Retrieved November 19, 2024, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063330/Green_Book_2022.pdf
- Joshanloo, M. (2025). A Person-Centered Exploration of Happiness Conceptions and Their Relation to Hedonic and Eudaimonic Well-Being. *Journal of Happiness Studies*, 26(7), 118. <https://doi.org/10.1007/s10902-025-00953-w>
- Kahneman, D. (1992). Reference points, anchors, norms and mixed feelings. *Organizational Behavior and Human Decision Processes*, 51, 296-312. [https://doi.org/10.1016/0749-5978\(92\)90015-Y](https://doi.org/10.1016/0749-5978(92)90015-Y)
- Kahneman, D., & Riis, J. (2005). Living, and thinking about it: two perspectives on life. In F. A. Huppert, N. Baylis, & B. Keverne, *The science of well-being* (pp. 285–304). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198567523.003.0011>
- Kaiser, C. (2022). Using memories to assess the intrapersonal comparability of wellbeing reports. *Journal of Economic Behavior & Organization*, 193, 410-442. <https://doi.org/10.1016/j.jebo.2021.11.009>
- Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences*, 119(42), e2210412119. <https://doi.org/10.1073/pnas.2210412119>
- Kaiser, C. & Lapinteur, A. (2025). Measuring the Unmeasurable? Systematic Evidence on Scale Transformations in Subjective Survey Data. *University of Oxford Wellbeing Research Centre*. Working Paper 2503. <https://doi.org/10.5287/ora-vyq7o5eqb>
- King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am. Political Sci. Rev.*, 98(1), 191–207. <https://www.jstor.org/stable/3593024>
- Layard, R. (2021). Wellbeing as the Goal of Policy. *LSE Public Policy Review*, 2(2), 1-8. <https://doi.org/10.31389/lseppr.46>
- Layard, R., & De Neve, J. E. (2023). *Wellbeing Science and Policy*. Cambridge University Press.

- Liberini, F., Redoano, M. and Proto, E. (2017) Happy voters. *Journal of Public Economics*, 146, p.41-57, <http://dx.doi.org/10.1016/j.jpubeco.2016.11.013>
- Linton, M-J., Dieppe, P., & Medina-Lara, A. (2016). Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. *BMJ open*, 6(7). <https://doi.org/10.1136/bmjopen-2015-010641>
- Liu, S., & Netzer, N. (2023). Happy Times: Measuring Happiness Using Response Times. *American Economic Review*, 113(12), 3289–3322. <https://doi.org/10.1257/aer.20211051>
- Loewenstein, G. (2009). That Which Makes Life Worthwhile. In Krueger, A. B. (org.) *Measuring the Subjective Well-Being of Nations*. National Bureau of Economic Research. pp.87-106.
- Lomas, T., Pawelski, J. O. & VanderWeele, T. J. (2024). A flexible map of flourishing: The dynamics and drivers of flourishing, well-being, health, and happiness. *International Journal of Wellbeing*, 13(4), 3665, 1-38. <https://doi.org/10.5502/ijw.v13i4.3665>
- Lyubomirsky, S. (2008). *The how of happiness: A scientific approach to getting the life you want*. Penguin Press.
- Mahoney, J. (2023). Subjective well-being measurement: Current practice and new frontiers. Working Paper No. 17. OEC Papers on Well-being and inequalities. <https://doi.org/10.1787/4e180f51-en>
- Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2021). Empirical relationships among five types of well-being. In: Matthew T. Lee, Laura D. Kubzansky, and Tyler J. VanderWeele (Eds). *Measuring well-being: Interdisciplinary perspectives from the social sciences and the humanities* (pp. 377-406). Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780197512531.003.0014>
- McClimans, L. (2013). The role of measurement in establishing evidence. *Journal of Medicine and Philosophy*, 38(5), 520-538. <https://doi.org/10.1093/jmp/jht041>
- McMahon, D. M. (2006). *Happiness: A history*. Grove Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O’Connon, K. J. (2023). Measuring Societal Progress. Working paper No.1. International Society for Quality of Life Studies. February 2023. <https://isqols.org/Workingpaperseries>
- Oishi, S., Graham, J., Kesebir, S., & Galinha, I. C. (2013). Concepts of happiness across time and cultures. *Personality and social psychology bulletin*, 39(5), 559-577. <https://doi.org/10.1177/0146167213480042>
- Panasiuk, S. L., McCanny, A., & Cheung, F. (2025). Methods reflect values: Evaluating the shortcomings of the average for measuring population well-being. *Journal of Personality and Social Psychology*, 128(6), p.1355-1370. <https://doi.org/10.1037/pspp0000549>
- Plant, M., McGuire, J., Dupret, S., Dwyer, R., & Stewart, B. (2025). Giving to others: how to convert your money into greater happiness for others. In Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J. E., Aknin, L. B., & Wang, S. World happiness report 2025. <https://doi.org/10.18724/whr-e0cy-0r69>
- Prati, A., & Senik, C. (2022). Feeling good is feeling better. *Psychological Science*, 33(11), 1828-1841. <https://doi.org/10.1177/09567976221096158>
- Prati, A., & Senik, C. (2025). Is it possible to raise national happiness? *Centre for Economic Performance*. Discussion Paper 2068. <https://cep.lse.ac.uk/pubs/download/dp2068.pdf>
- Rodriguez Duque, S., Tal, E., & Barbic, S. P. (2024). The role of ethical and social values in psychosocial measurement. *Measurement*, 225, 113993. <https://doi.org/10.1016/j.measurement.2023.113993>
- Samuelsson, C., Dupret, S., Plant, M., & Kaiser, C. (2023). Can we trust wellbeing surveys? A pilot study of comparability, linearity, and neutrality. Happier Lives Institute. Retrieved November 19, 2024, from <https://www.happierlivesinstitute.org/comparability-linearity-and-neutrality-pilot-report/>
- Schneider, L., & Schimmack, U. (2009). Self-informant agreement in well-being ratings: A meta-analysis. *Social Indicators Research*, 94(3), 363–376. <https://doi.org/10.1007/s11205-009-9440-y>
- Schutte, L., Negri, L., Delle Fave, A. & Wissing, M. P. (2021). Rasch analysis of the satisfaction with life scale across countries: Findings from South Africa and Italy. *Current Psychology*, 40(10), 4908–4917. <https://doi.org/10.1007/s12144-019-00424-5>
- Sewaybricker, L. E., & Massola, G. M. (2022). Against well-being: A critique of positive psychology. *History of the Human Sciences*, 36(1), 131-148. <https://doi.org/10.1177/095269512211147>
- Sewaybricker, L., & Massola, G. M. (2025). Dialogues on happiness with the Guarani Mbyá in South America and the limits of the science of well-being. *International Journal of Wellbeing*, 15(2). <https://doi.org/10.5502/ijw.v15i2.2759>

- Sollis, K., Yap, M., Campbell, P., & Biddle, N. (2022). Conceptualisations of wellbeing and quality of life: A systematic review of participatory studies. *World Development*, *160*, 1-39. <https://doi.org/10.1016/j.worlddev.2022.106073>
- Stevenson, B., & Wolfers, J. (2008). Economic growth and subjective well-being: Reassessing the Easterlin paradox (No. w14282). *National Bureau of Economic Research*. <https://doi.org/10.3386/w14282>
- Tal, E. (2020). *Measurement in Science*. The Stanford Encyclopaedia of Philosophy. Retrieved November 19, 2024, from <https://plato.stanford.edu/archives/fall2020/entries/measurement-science>
- Veenhoven, R. (1984). *Conditions of Happiness*. Dordrecht: Springer. <http://dx.doi.org/10.1007/978-94-009-6432-7>
- Veenhoven, R. (2017). Measures of Happiness: Which to Choose? In G. Brulé, & F. Maggino (Eds.), *Metrics of Subjective Well-Being: Limits and Improvements* (pp. 65-84). Springer International Publishing. https://doi.org/10.1007/978-3-319-61810-4_4
- Veenhoven, R. (2025). *World Database of Happiness*. Erasmus University Rotterdam, The Netherlands. Retrieved November 19, 2024, from <https://worlddatabaseofhappiness.eur.nl>
- Ward, G., De Neve, J.-E., Ungar, L. H., & Eichstaedt, J. C. (2021). (Un)happiness and voting in U.S. presidential elections. *Journal of Personality and Social Psychology*, *120*(2), 370–383. <https://doi.org/10.1037/pspi0000249>
- Weijers, D., & Jarden, A. (2013). The science of happiness for policymakers: An overview. *Journal of Social Research & Policy*, *4*(2), 21 - 40. <https://repository.openpolytechnic.ac.nz/handle/11072/1629>
- Weiss, S., & Roberts, R. D. (2018). Using anchoring vignettes to adjust self-reported personality: A comparison between countries. *Frontiers in psychology*, *9*, 325. <https://doi.org/10.3389/fpsyg.2018.00325>
- Wierzbicka, A. (2004) "'Happiness" in Cross-Linguistic & Cross-Cultural Perspective', *Daedalus*, *133*(2), 34-43. <https://doi.org/10.1162/001152604323049370>
- White, N. (2006). *A brief history of happiness*. New Jersey, NY: Blackwell Publishing. <https://doi.org/10.1002/9780470690888>
- Yap, M., & Yu, E. (2016). Operationalising the capability approach: developing culturally relevant indicators of indigenous wellbeing – an Australian example. *Oxford Development Studies*, *44*(3), 315–331.