Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparing Countries Is Not Following Them: When Can a Cross-National Well-Being Decomposition Describe Change?

Shay Tsaban

shay.tsaban@gmail.com

Peres Academic Center

Research Article

Keywords: World Happiness Report, life evaluation, progress dashboards, between and within variation, well-being indicators, data vintages

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11554706/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Comparing Countries Is Not Following Them: When Can a Cross-National Well-Being Decomposition Describe Change?

Abstract

Subjective well-being is increasingly proposed as a headline indicator for tracking progress. The World Happiness Report divides each country's life evaluation into parts attributed to six factors, using coefficients estimated across countries, and the parts are often read as explaining why a country's well-being has changed. This paper asks when that reading is justified and states three conditions. The factors must relate to life evaluation within countries as they do between them (structural equivalence), the relationships must be stable over time (temporal invariance), and the published series must come from one model and one set of data throughout (measurement continuity). A probe with the report's public files for 2019 to 2025 shows that the third condition fails by construction. Each year of the published series comes from a different edition, with its own coefficients, benchmarks and data sources, and the income and health series break between editions. Rebuilt on one set of coefficients and consistent income and life expectancy data, the series shows within-country slopes for income and social support at least as large as their pooled cross-country slopes, although the income difference is not statistically significant, while the within-country slope for freedom is a quarter to a third of the pooled slope. Within-country slopes also differ between the two halves of the period. The decomposition is a defensible account of differences between countries but, as published, not an account of change. The paper proposes reporting practices, including for progress dashboards, that would support a dynamic reading.

Keywords World Happiness Report; life evaluation; progress dashboards; between and within variation; well-being indicators; data vintages

Introduction

Subjective well-being has become a leading candidate among measures of progress beyond gross domestic product. Recent work in this journal argues that it should serve as the headline indicator of progress dashboards, with measures of the conditions that support it reported alongside (Burger et al., 2026), and that happiness, understood as life satisfaction, is the most inclusive indicator of quality of life in regions and can be measured and compared across place and time (Veenhoven, 2026). The updated OECD guidelines set out how national statistical offices should measure it (OECD, 2025), new instruments aim to follow population happiness in close to real time (Greyling & Rossouw, 2025), and governments that adopt well-being as an objective of policy need to know what moves their own population's score (Frijters et al., 2020). All of these uses are dynamic. They ask whether well-being in a given place is rising or falling, and why.

The best-known instrument that already combines a headline measure with the conditions that support it is the World Happiness Report. Each year it ranks countries by the average answer of their residents to the Cantril ladder question and divides each country's score into parts attributed to income, social support, healthy life expectancy, freedom to make life choices, generosity and perceptions of corruption (Helliwell et al., 2026). The coefficients behind that division are estimated from differences between countries. They answer a comparative question: why are some countries happier than others? Readers, journalists and governments routinely use the published parts to answer a dynamic question: why did this country become happier or less happy? This paper asks when an answer to the first question can serve as an answer to the second.

Parts of the answer are familiar. The Easterlin paradox is a finding about exactly this gap for one factor: income differences track happiness differences across countries, while income growth within a country tracks national happiness much more weakly over the long run

(Easterlin, 1974; Easterlin et al., 2010; Easterlin & O'Connor, 2022). The conclusion is contested by work that finds similar slopes in both comparisons (Stevenson & Wolfers, 2008; Sacks et al., 2012) and by country-level panel evidence that finds no paradox once income inequality is taken into account (Smith & Wesselbaum, 2023). Econometrics has long distinguished the between-unit and within-unit components of a pooled coefficient (Mundlak, 1978; Bell & Jones, 2015), and psychology has a parallel literature on separating between-person from within-person effects (Curran & Bauer, 2011). Quality-of-life research adds evidence that bears on attribution: the relationship between income inequality and life satisfaction differs across countries (Schilpzand & Graafland, 2026), national well-being profiles persist across survey waves (Hill, 2026), and national happiness does not reliably return to a stable set point after shocks (Mueller, 2023). Much of what distinguishes countries may therefore be stable features of the countries, while what moves within them may follow its own dynamics.

What is missing is an account of the conditions under which a published cross-national attribution may be read as a description of change, and a check of whether a working instrument meets them. If dashboards are to place subjective well-being at their head and report its supporting conditions below it (Burger et al., 2026), users will read the supporting conditions as an explanation of movement in the headline. Studies that identify the predictors of well-being across countries, including recent machine-learning approaches (Li & Hu, 2026), address the comparative question. Research on composite indicators has shown that rankings and attributions can be fragile to choices rarely visible to their users (Saisana et al., 2005; Ravallion, 2012; Høyland et al., 2012; Greco et al., 2019). Neither literature says when the comparative answer may be used for the dynamic question.

The aim of this paper is to provide that account and that check. Its contributions are threefold. First, it states three conditions for transferring an attribution from comparison to

change: structural equivalence of between-country and within-country relationships, temporal invariance of those relationships, and measurement continuity of the published series. The first two organize existing arguments in the between-within literature. The third is new to well-being research, although it is standard in work on data revisions in macroeconomics (Croushore & Stark, 2001; Johnson et al., 2013), and in this application it is the one that fails first. Second, it shows that the World Happiness Report's published country-year decomposition fails measurement continuity by construction. Each year in the series is the decomposition published in the following year's edition, with different coefficients, different reference values and, for income and health, different data sources. This alone produces the appearance that income and healthy life expectancy carry no information about a country's movement, a conclusion that an analysis of the published series would reach and that, for income, does not survive once the series is rebuilt. Third, it probes the two substantive conditions directly, using a series rebuilt on a single set of coefficients and on consistent external data. Structural equivalence cannot be rejected for income, social support, life expectancy, generosity and perceptions of corruption, but it fails for freedom, and within-country relationships shift between the two halves of a six-year period. The quantitative part is a plausibility probe: it shows that each condition can fail in a real instrument and which ones do, and it does not estimate causal effects.

The rest of the paper is organized as follows. The section "One Instrument, Two Questions" describes the decomposition and the two questions it can be asked. The "Conceptual Framework" section develops the three conditions. "Data and Empirical Strategy" sets out the design of the probe, and "Results" reports what it shows. The "Discussion" draws implications for reporting practice, for progress dashboards and for research on change in national well-being, and the last section concludes.

One Instrument, Two Questions

The Comparative Question and the Dynamic Question

A national well-being statistic can be asked two kinds of question. The comparative question asks why the average life evaluation in one country differs from that in another at a given time. The dynamic question asks why the average in a given country has risen or fallen. The two are often run together because the same variables appear in both answers, but they concern different sources of variation. The comparative answer draws on everything that distinguishes countries, including features that change slowly or not at all: institutions, history, climate, language, social norms and the conventions people follow when they answer survey questions. The dynamic answer draws only on what changes within a country over the period considered.

An attribution makes this difference sharper. A statistic that reports a level makes no claim about causes. A statistic that divides the level into parts attached to named factors makes a claim that each part is owed to its factor. That claim has to be read with respect to some comparison, because a regression coefficient describes how the outcome differs across the variation from which it was estimated. If the coefficient is estimated from differences between countries, the attribution says how much of a country's distance from a reference country is associated with each factor. It does not by itself say how much of a country's change is associated with each factor.

Validity Belongs to a Use

Measurement theory treats validity as a property of an interpretation or use of a score rather than of the score itself (Messick, 1995; Kane, 2013). A test can support one inference and not another, and each intended use calls for its own evidence. Applied to national well-being, this means that the question is not whether the World Happiness Report's decomposition is valid but whether it is valid for a given use. Evidence that the

decomposition orders countries well is evidence for the comparative use. It is not evidence for the dynamic use, which needs its own argument and its own checks. The same distinction applies to the claim that happiness is comparable across place and time (Veenhoven, 2026). That claim concerns the measure. Whether an attribution of the measure to its conditions is comparable across time is a separate question.

The literature on composite indicators reaches a similar point from a different direction. Composite and decomposed indicators acquire authority from their form, and the choices embedded in them are rarely visible to their audiences (Ravallion, 2012; Høyland et al., 2012). Rankings built from them can be less robust than their presentation implies (Saisana et al., 2005; Paruolo et al., 2013; Greco et al., 2019). The present case differs from the usual composite indicator problem in one respect. The score being decomposed exists independently of the decomposition, since the ranking uses the survey average alone. The question is therefore not how parts are aggregated into a score but how an already measured score is attributed to causes, and for which comparison the attribution holds.

The Decomposition and the Form in Which It Is Published

The report's ranking uses the national mean of the Cantril ladder, a zero to ten scale on which respondents place their present life, averaged over the three most recent survey years of the Gallup World Poll. The decomposition is a separate exercise. National annual means of life evaluation are regressed on six national characteristics, pooled across countries and years with year fixed effects and standard errors clustered by country (Helliwell et al., 2026, Statistical Appendix, Table 8). For each country, the contribution of a factor is the estimated coefficient multiplied by the distance between the country's three-year average of that factor and the value assigned to a hypothetical reference country, Dystopia, which has the lowest observed value of each factor. What remains after the six contributions are subtracted is

published as Dystopia plus residual. The six contributions and the remainder add up to the published life evaluation.

Within any one edition, every country is decomposed with the same coefficients and the same reference values, so the edition's decomposition is a consistent account of differences between countries in that year. The report also publishes the decomposition as a country-year series. The data file that accompanies the 2026 edition lists, for each country, the six contributions and the remainder in every year from 2019 to 2025. A reader who follows one country's bar across these years is comparing the country with itself. The segments appear to show how much of the country's score, and by extension of any change in it, is owed to each factor. Nothing in the file or the figure indicates whether the segment widths in different years were computed with the same coefficients and the same data.

The decomposition has received methodological attention before. Tofallis (2020) questions the additive form and proposes a multiplicative alternative. Nilsson et al. (2024) show that the ladder wording steers respondents toward thoughts of power and wealth. Padgett et al. (2026) find that the ladder is less strongly correlated with other dimensions of flourishing than a life satisfaction question in all 22 countries they study. The comparability of ladder responses across countries and the cardinal treatment of the scale remain debated (Bjørnskov, 2010; Bond & Lang, 2019; Kaiser & Oswald, 2022), as does the measurement invariance of subjective well-being across countries (Röck et al., 2026). These critiques concern how the decomposition is built and what the outcome measures. The question here is logically prior to them. Taking the published formula and outcome as given, it asks which comparison the published contributions are entitled to describe.

Conceptual Framework: When Can a Between-Country Attribution Describe Change?

A Model with Both Sources of Variation

Let the average life evaluation in country c in year t be

$$h_{ct} = \mathbf{X}_{ct}\mathbf{B}_c + \lambda_t + \alpha_c + u_{ct}, \qquad (1)$$

where $\mathbf{X}_{ct}$ is the vector of the six factors, $\mathbf{B}_c$ is a vector of slopes that may differ across countries, λ_t is a year effect common to all countries, α_c is a persistent country effect and u_{ct} collects everything else. The report estimates a single vector **B** by pooled least squares with year effects, omitting α_c. That estimate mixes two quantities: the relationship between factors and life evaluation across countries, and the relationship within countries over time (Mundlak, 1978; Bell & Jones, 2015). The pooled coefficient is a weighted combination of a between-country coefficient and a within-country coefficient, with weights set by how much of each factor's variance lies between and within countries. For national characteristics most of the variance lies between countries, so the pooled coefficient is close to the between-country one.[1]

Reading the published contributions as an account of a country's change means treating the change in a contribution, the pooled coefficient times the change in the factor, as the part of the change in life evaluation associated with that factor. If the within-country slope differs from the pooled one, the attributed change is off by the difference between the two slopes times the change in the factor. The error is therefore largest for exactly the factors and countries that moved most, which are the cases a dynamic reading is most interested in. It does not shrink as more countries are added, because it is a property of the comparison the coefficient describes, not of sampling. Three conditions are needed for the error to vanish, and each can fail independently (Fig. 1).

[1] In the probe below, the between-country share of the variance of each factor over 2019 to 2025 exceeds 92 per cent for every factor; see Online Resource 1, Section S3.

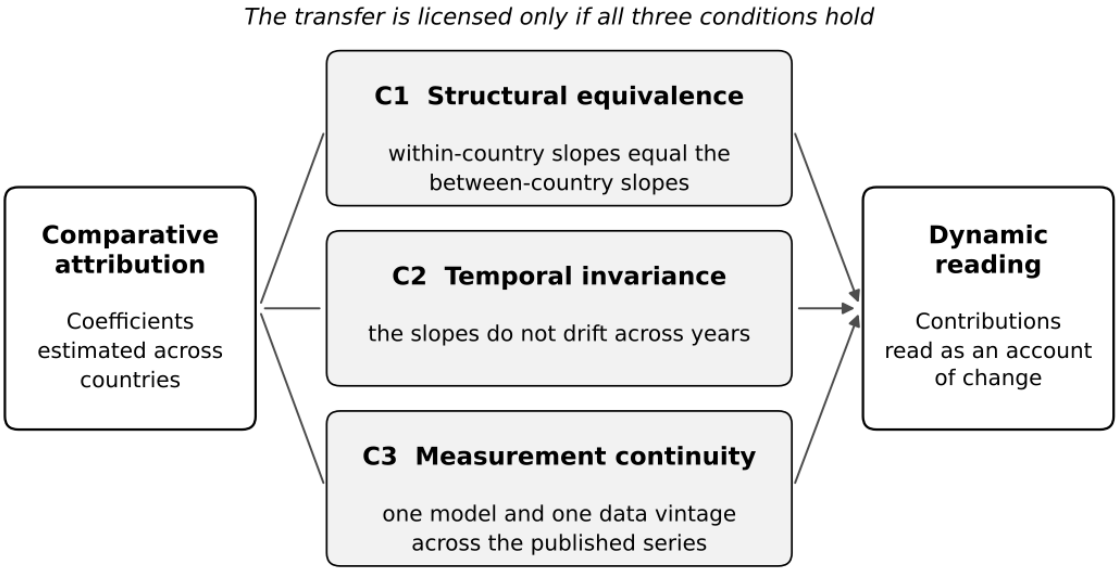


Fig. 1 Conditions under which a between-country attribution can be read as an account of change within a country

Condition 1: Structural Equivalence

The first condition is that the within-country relationship equals the between-country relationship, and that it is common to countries, so that $\mathbf{B}_c = \mathbf{B}$ and the persistent country effect α_c is unrelated to the factors. There are four reasons to expect the condition to fail, and they do not all point in the same direction.

The first is stable confounding. Countries that differ in income also differ in institutions, history, culture and survey response styles, and these differences are correlated with most of the six factors (Exton et al., 2015). National well-being profiles are largely stable across two waves of a 22-country survey and are not captured by income alone (Hill, 2026), which is what persistent country effects would produce. A between-country coefficient absorbs them; a within-country coefficient does not. The second is adaptation and comparison. If people adapt to improved circumstances or judge their lives relative to others, a lasting rise in a factor shared by everyone in a country can leave average life evaluation unchanged even though people in richer or healthier countries report higher life evaluations (Clark et al., 2008; Luhmann et al., 2012). Relative effects need not run through social comparison and

aspirations alone; they can also work through the capabilities that a given level of a resource affords relative to others (Hirata & Groß, 2025). This argument predicts within-country slopes smaller than between-country slopes, especially for income over long horizons, and it is the core of Easterlin's interpretation of his paradox (Easterlin & O'Connor, 2022). The third is horizon. Over short periods, national income moves with the business cycle, and recessions lower life evaluation more than expansions raise it (Di Tella et al., 2003; De Neve et al., 2018). Short-run within-country slopes can therefore be larger than long-run slopes, and even larger than between-country slopes. The fourth is measurement error. Within-country comparisons remove the persistent part of each variable and leave a larger share of noise, which biases within-country slopes toward zero (Griliches & Hausman, 1986).

The factors are likely to differ in how far these forces apply. Social support is a relational resource that can change quickly with circumstances, and there is evidence that it moves with life evaluation over time as well as across places (Helliwell et al., 2014; Bartolini & Sarracino, 2014), including individual fixed effects evidence that the number of friends has a small positive effect on life satisfaction (Mader & Franzen, 2025). Freedom to make life choices has been proposed as a driver of rising happiness (Inglehart et al., 2008), but its association with life evaluation across countries may also carry slowly changing features of political and economic institutions that a within-country comparison removes. Cross-national studies find that life satisfaction differs with regime type in a nonlinear way (Prati, 2022) and that the rule of law is related to happiness, mostly directly rather than through civic virtues and trust (Graafland, 2023), and these are features that change slowly. Consistent with this, an analysis of 165 countries finds no clear overall link between autocratization and life satisfaction, and a negative one for several democracy measures when established democracies become less democratic (Bromo & Pacek, 2026). Slope heterogeneity is a separate concern. The relationship between income and life satisfaction varies across

countries with inequality, social trust and wealth (Oishi & Kesebir, 2015; Mikucka et al., 2017; Brzezinski, 2025), the relationship between inequality and life satisfaction varies with income per capita and national culture (Schilpzand & Graafland, 2026), and country-specific measures of subjective well-being capability can deviate substantially from conventional well-being measures even where a general cross-country model tracks them (Kwarciński et al., 2024). A single pooled slope describes individual countries only on average.

Condition 2: Temporal Invariance

The second condition is that the relationships do not change over the period in which the contributions are compared. A contribution computed with this year's coefficient and compared with one computed with last year's coefficient describes change in the factor only if the underlying relationship has stayed the same. There are reasons to expect drift. Crises can change the salience of particular resources; social trust and support appear to matter more for life evaluation during them (Helliwell et al., 2014), and life evaluation responds to events and crises in ways that stable characteristics do not anticipate (Kapteyn et al., 2025; Kriechel et al., 2025). National happiness does not reliably return to a fixed set point either: in more than half of the European countries studied by Mueller (2023) no homeostasis is found, and where it is, the set point often shifts or holds only within a narrow range. The pandemic is the obvious recent case of a shock: losses of income and work and their relation to well-being have been examined directly with the Gallup World Poll (Ochoa Pacheco et al., 2026). The set of countries surveyed also changes from year to year, which alters what a pooled coefficient averages over.

Condition 3: Measurement Continuity

The third condition is that the published series is built from one model and one set of data throughout. It is usually taken for granted, and the between-within literature does not discuss it. It is nevertheless a precondition for any dynamic reading, because a difference

between two published contributions measures change in the factor only if both were computed with the same coefficient, the same reference value and the same data source.

Three kinds of vintage can differ across years of a published series. The model vintage is the set of coefficients. The benchmark vintage is the reference country against which contributions are measured. The data vintage is the release of each underlying variable. Macroeconomics has a large literature on the last of these. Data revisions alter the measured path of output (Croushore & Stark, 2001), successive versions of the Penn World Table change estimated growth rates enough to overturn published findings (Ponomareva & Katayama, 2010; Johnson et al., 2013), and new rounds of the International Comparison Program revise purchasing power parities substantially (Deaton & Aten, 2017). A series that splices values taken from successive releases records revisions as if they were changes. Official statistics avoid this by publishing back-cast series in which every year is recomputed on the current vintage. A well-being decomposition published as a country-year series raises the same issue, with the added complication that two of its six factors, income and healthy life expectancy, are drawn from administrative sources that are revised between releases.

The consequences can be written out. Let a factor's published contribution in year t be the coefficient of the edition used for year t times the distance of the country's value from that edition's reference value. The change in the contribution between two years then has three parts: the later coefficient times the change in the factor; the change in the coefficient times the country's earlier distance from the reference value; and the later coefficient times the change in the reference value. The third part is the same for every country and drops out of any comparison of change across countries. The second does not. It is proportional to how far a country stood from the reference country at the start, so it is correlated with the country's initial position on the factor and, through it, with its initial life evaluation. A revision of a coefficient therefore creates an apparent change that is systematically larger, in absolute

terms, for countries that started high, and it can mimic or mask regression toward the mean in the outcome. A data revision enters through the first part, as change in the factor that did not happen. Neither is visible in a stacked bar.

Measurement continuity has a property that the other two conditions lack. It can be checked from the published files without any assumption about causal structure, and if it fails, the published series cannot support a dynamic reading whatever is true of the other two conditions.

What the Conditions Imply

The three conditions can be checked in order. If measurement continuity fails, the published contributions cannot be compared over time and must be rebuilt before anything else can be said. If the series can be rebuilt, structural equivalence can be tested by comparing pooled and within-country slopes, and temporal invariance by comparing slopes across periods. The framework predicts no single pattern across factors. Stable confounding and adaptation predict within-country slopes smaller than between-country slopes. Short-horizon cyclical effects predict the opposite for income. Relational factors such as social support are the most plausible candidates for transfer. The probe that follows is designed to show whether these possibilities are realized in the most widely circulated decomposition of national well-being.

Data and Empirical Strategy

Data

The probe uses the public data file published with the World Happiness Report 2026, which contains life evaluations, their 95 per cent confidence intervals, the six contributions and the remainder for country-years from 2019 to 2025 (Helliwell et al., 2026; World

Happiness Report, 2026).[2] After country-years with any missing term are dropped, the panel contains 1,013 country-years for 157 countries. The six contributions and the remainder add up to the published life evaluation within 0.003 ladder points, which is consistent with rounding in the file. To audit the series, the probe also uses the corresponding files from the 2020, 2021, 2023 and 2025 editions and the regression tables in each edition's statistical appendix. Income and longevity are also measured from a single release of the World Bank's World Development Indicators: gross domestic product per capita in purchasing power parity terms at constant 2021 prices, in logarithms, and life expectancy at birth (World Bank, 2026). Both are averaged over the same three survey years as the life evaluation they are compared with.[3]

Rebuilding a Single-Vintage Series

Each published contribution equals an edition-specific coefficient multiplied by the distance between the country's value of the factor and the edition's reference value. Dividing a contribution by its edition's coefficient therefore recovers the factor itself, measured from that edition's reference value. Within a year all countries share the same reference value, so year effects absorb it, and the recovered factors can be compared across years without knowing the reference values. Multiplying the recovered factors by one coefficient vector, here that of the 2026 edition, gives a harmonized decomposition in which only the data change from year to year. Correlations and slopes for individual factors do not depend on which edition's coefficients are used for harmonization, since a common rescaling leaves them unchanged; the choice matters only for the summed contribution. For the four survey-based factors harmonization removes the model and benchmark vintages. For income and

[2]The file lists life evaluations from 2011 onward; the decomposition is published only from 2019. Online Resource 1, Table S1, lists every file with its source; Online Resource 3 records checksums.

[3]For the 2025 wave, life expectancy averages 2023 and 2024, the latest years available in the release used. Taiwan and North Cyprus are not covered by the World Development Indicators and drop out of analyses that use them.

health it does not remove the data vintage, which is why the probe replaces them with the World Bank series.

Estimation and Inference

Structural equivalence is examined by estimating equation (1) in two ways: pooled least squares with year effects, which is the report's specification applied to the probe's panel, and least squares with country and year effects, which identifies the slopes from movement within countries only. The difference between the two is tested with the correlated random effects form of the Mundlak (1978) model, in which country means of the six factors and of the year indicators are added to the pooled regression; their coefficients estimate the difference between the between-country and within-country slopes, and a joint test that they are zero is a test of the pooled model against the fixed effects model (Hausman, 1978; Bell & Jones, 2015). Temporal invariance is examined by testing whether cross-sectional slopes differ across years and whether within-country slopes differ between the two adjacent windows 2019 to 2022 and 2022 to 2025.

The published life evaluations are three-year moving averages, so adjacent years share survey data. The preferred specifications therefore use three waves, 2019, 2022 and 2025, which summarize the survey years 2017 to 2019, 2020 to 2022 and 2023 to 2025 and share no survey year. All seven years are used as a check. Standard errors are clustered by country. Because cluster-robust Wald tests with many restrictions and about 150 clusters can over-reject, joint tests are also evaluated with a restricted wild cluster bootstrap (Cameron et al., 2008; Cameron & Miller, 2015). The countries analysed are close to the full set covered by the report, but sampling error in the survey means and omitted country-year influences make the slopes uncertain, so standard errors and tests are reported and interpreted in the usual model-based sense (Abadie et al., 2020). The estimates describe associations in the published data. They are not estimates of causal effects, and the probe is designed to test the conditions

under which the published attribution can be read in a particular way, not to identify what moves national well-being.

Results

The Published Series Fails Measurement Continuity

Each year of the series in the 2026 file is the decomposition published in the following year's edition. For 2019, 2020 and 2022, the years whose original files report the underlying factors, the contributions in the 2026 file match the original edition's contributions to within rounding (maximum absolute difference 0.0005 ladder points), and in each original file every contribution is an exact linear function of the factor, which identifies the coefficient used. The 2025 file repeats the same values for all 868 overlapping country-years, which confirms the pattern for 2024 and shows that a comparison between the two files cannot reveal the splice. For 2021 and 2023, whose original files do not report the factors, the probe assumes the same structure and applies the coefficients in those editions' appendices.

Table 1 Coefficients used for each year of the published country-year decomposition

Data year	Edition	Income	Social support	Healthy life expectancy	Freedom	Generosity	Corruption
2019	WHR 2020	0.310	2.362	0.036	1.199	0.661	−0.646
2020	WHR 2021	0.349	2.253	0.031	1.217	0.652	−0.638
2021	WHR 2022	0.360	2.420	0.029	1.305	0.583	−0.704
2022	WHR 2023	0.359	2.526	0.027	1.331	0.537	−0.716
2023	WHR 2024	0.349	2.563	0.028	1.378	0.487	−0.733
2024	WHR 2025	0.328	2.686	0.032	1.518	0.382	−0.669
2025	WHR 2026	0.297	2.820	0.035	1.546	0.339	−0.662

Each year of the series in the World Happiness Report 2026 data file is the decomposition published in the edition named. Coefficients for 2019, 2020 and 2022 are those implied exactly by the original edition files (for 2020 they equal Table 11, column 1, of the 2021 appendix); the others are from Table 8, column 1, of each edition's statistical appendix. Units are ladder points per unit of the factor. Sources and page references are listed in Online Resource 1, Section S2.

The coefficients differ across editions (Table 1). The coefficient applied to social support rises from 2.36 for 2019 to 2.82 for 2025, and that applied to freedom from 1.20 to 1.55. The coefficient applied to generosity halves, from 0.66 to 0.34 (Fig. 2a). The reference values also

change. The Dystopia value of the logarithm of income is 6.49 in the 2020 edition and 5.53 in the 2023 edition; that of healthy life expectancy rises from 45.2 to 51.5 years. Changes in reference values shift every country's contribution by the same amount within a year and do not affect comparisons of change across countries. Changes in coefficients do, because they rescale each country's contribution in proportion to its distance from the reference value. For generosity, the change in the coefficient alone accounts for about 66 per cent of the cross-country variance of the change in the published contribution between 2019 and 2025.[4]

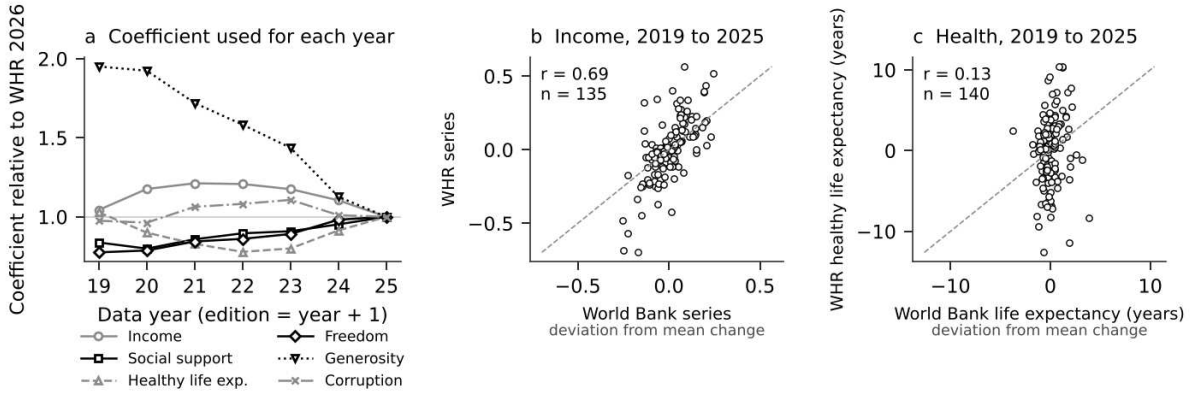


Fig. 2 The published country-year series splices edition vintages. (a) Coefficient applied to each data year, relative to the 2026 edition. (b) Change in income per capita, 2019 to 2025: the report's series against the World Bank series. (c) Change in healthy life expectancy in the report against change in World Bank life expectancy. Panels b and c show deviations from the cross-country mean change; the dashed line marks equality

The data vintages matter more for income and health. Between the 2020 and 2021 editions the median country's logarithm of income per capita rises by 0.108, while the World Bank series for the same windows shows a median change of −0.002. A rise of this size over a window that ends in 2020 is most plausibly a change in the price base of the income series, not growth. Between the 2021 and 2023 editions healthy life expectancy changes with a cross-country standard deviation of 1.8 years, against 0.7 years for life expectancy in the World Bank series, and a further break appears between the 2024 and 2025 editions, where

[4]The change in a published contribution between 2019 and 2025 equals the 2025 coefficient times the change in the factor plus the change in the coefficient times the 2019 value of the factor. Online Resource 1, Table S4, reports the variance of each part for every factor.

the correlation of healthy life expectancy levels across adjacent years falls to 0.91. Over the whole period, the change in the report's income series correlates 0.69 with the change in the consistent World Bank series, and the change in the report's healthy life expectancy correlates 0.13 with the change in World Bank life expectancy (Fig. 2b and 2c). For health, the published change is almost entirely revision.

Table 2 Correlation of each factor with life evaluation between countries and within countries

Factor	Between countries, 2025	Within, 2019 to 2025: published	Within, 2019 to 2025: one coefficient	Within, 2019 to 2025: consistent data	Within, 2022 to 2025: published	Within, 2022 to 2025: consistent data
Log GDP per capita	0.75	−0.09	−0.07	0.27	0.02	0.45
Social support	0.81	0.54	0.54	n/a	0.48	n/a
Health	0.68	−0.18	−0.18	−0.10	−0.18	0.24
Freedom to make life choices	0.64	0.38	0.34	n/a	0.23	n/a
Generosity	0.04	0.23	0.07	n/a	−0.02	n/a
Perceptions of corruption	0.40	0.19	0.20	n/a	0.18	n/a

Between countries: correlation across countries of the published contribution with life evaluation in 2025. Within: correlation across countries of the change in the factor with the change in life evaluation. Published: change in the published contribution. One coefficient: recomputed with the 2026 coefficients for every year. Consistent data: World Bank series (log GDP per capita at constant 2021 PPP; life expectancy for health). n/a: no external series used.

The consequences for a dynamic reading are shown in Table 2. In the published series the change in the income contribution between 2019 and 2025 is unrelated to the change in life evaluation (correlation −0.09) and the change in the healthy life expectancy contribution is negatively related to it (−0.18). With income measured consistently, the correlation is 0.27 over the whole period and 0.45 between 2022 and 2025. The apparent absence of any within-country income signal in the published decomposition is a property of how the series is assembled, not of how income relates to life evaluation. For health the consistent series does not reverse the conclusion: life expectancy is only weakly related to change in life evaluation (−0.10 over the whole period, 0.24 between 2022 and 2025). For generosity the published

correlation of 0.23 falls to 0.07 once one coefficient is applied to all years. Social support and freedom are barely affected by harmonization.

Structural Equivalence Holds for Some Factors and Fails for Freedom

Table 3 compares pooled and within-country slopes in the harmonized panel, with income and life expectancy from the World Bank series. Fig. 3 puts the two sets of slopes on a common scale, the change in life evaluation associated with a difference of one cross-country standard deviation in each factor.

Table 3 Between-country and within-country slopes of life evaluation on the six factors, three non-overlapping waves

Factor	Between countries (pooled)	Within countries (fixed effects)	Difference, between minus within	p value	Ratio, within to pooled
Log GDP per capita (World Bank)	0.27 (0.09)	0.73 (0.32)	−0.45 (0.34)	0.190	2.66
Social support	3.47 (0.55)	4.17 (0.56)	−0.83 (0.89)	0.348	1.20
Life expectancy (World Bank)	0.013 (0.013)	0.036 (0.019)	−0.024 (0.023)	0.278	2.73
Freedom to make life choices	2.16 (0.40)	0.55 (0.41)	1.77 (0.63)	0.005	0.25
Generosity	0.15 (0.29)	0.38 (0.30)	−0.25 (0.44)	0.580	2.44
Perceptions of corruption	−0.69 (0.25)	−1.01 (0.38)	0.31 (0.48)	0.518	1.46
Joint test of equal slopes			$\chi^2(6) = 10.16$	0.118	bootstrap p = 0.200
Sample	152 countries, 421 observations	139 countries with two or more waves identify the slopes			

Waves 2019, 2022 and 2025 (survey years 2017 to 2019, 2020 to 2022 and 2023 to 2025). Pooled: least squares with year effects. Within: least squares with country and year effects. Difference: coefficient on the country mean of the factor in the correlated random effects (Mundlak) model, which includes country means of the factors and of the year indicators. Standard errors clustered by country in parentheses. Bootstrap: restricted wild cluster bootstrap, 1,999 replications. Results for all seven years are in Online Resource 1, Table S7.

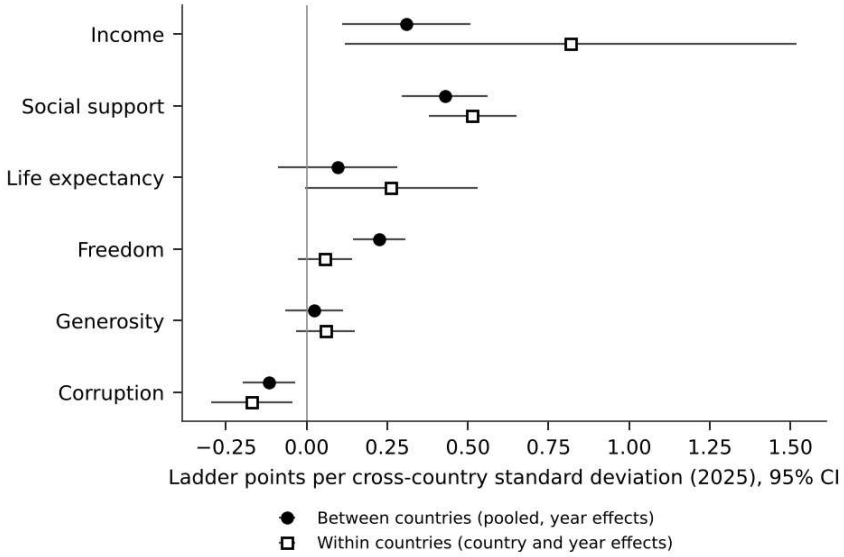


Fig. 3 Between-country and within-country slopes on a common scale. Slopes from Table 3 multiplied by the cross-country standard deviation of each factor in 2025, with 95 per cent confidence intervals

The pattern matches none of the simple expectations. For social support the within-country slope (4.17) is somewhat above the pooled slope (3.47), and the difference is not statistically distinguishable from zero ($p = 0.348$). Social support transfers. For income the within-country slope (0.73, standard error 0.32) is larger than the pooled slope (0.27), although it is less precisely estimated and the difference is not statistically distinguishable from zero ($p = 0.190$). A within-country slope above the cross-country slope over a six-year window that contains a global recession and a recovery is what the horizon argument set out under structural equivalence predicts, and it is consistent with evidence that short-run income fluctuations move national life evaluations (Di Tella et al., 2003; De Neve et al., 2018). It is not evidence against the long-run version of the Easterlin paradox, which concerns trends rather than fluctuations. Perceptions of corruption and generosity show within-country slopes of the same sign as, and not distinguishable from, the pooled slopes. Life expectancy has positive slopes in both comparisons once the other factors are held constant, but both are imprecisely estimated.

Freedom is the exception. Its within-country slope (0.55, standard error 0.41) is about a quarter of its pooled slope (2.16), and the difference is clearly distinguishable from zero ($p = 0.005$; wild bootstrap $p = 0.006$). The annual panel gives a ratio of about a third. A country's position on freedom accounts for a large part of its distance from other countries, but changes in freedom within a country over these years are associated with much smaller changes in life evaluation. This is the pattern that stable confounding predicts: differences in freedom between countries carry institutional and cultural differences that a within-country comparison holds fixed.

Taken together, the six differences give little joint evidence against the pooled model. The joint Mundlak test gives $\chi^2(6) = 10.16$ ($p = 0.118$) in the three-wave panel and $p = 0.097$ in the annual panel; the wild bootstrap gives $p = 0.200$ and $p = 0.204$. The evidence against the pooled slopes is concentrated in freedom, whose difference is distinguishable on its own in both panels (annual panel, wild bootstrap $p = 0.005$). The report's own statistical appendix points in the same direction. Its specification with country fixed effects, estimated on the report's consistent annual data from 2005, gives a larger income coefficient than the pooled specification (0.414 against 0.297), a smaller freedom coefficient (1.107 against 1.546) and a similar social support coefficient (2.559 against 2.820) (Helliwell et al., 2026, Statistical Appendix, Tables 8 and 10). Those fixed effects estimates are not used in the published decomposition.

Temporal Invariance Is Doubtful

The relationships also move over time. Within-country slopes estimated separately for 2019 to 2022 and 2022 to 2025 differ jointly (wild bootstrap $p = 0.008$). The income slope is 0.19 (standard error 0.43) in the first window, which spans the pandemic, and 1.69 (0.40) in the second; the social support slope falls from 4.76 to 2.51. Cross-sectional slopes drift less clearly. Estimated year by year, the social support slope rises from 2.76 in 2019 to 3.89 in

2025 and the generosity slope falls from 0.38 to −0.05, but a joint test of equality across the three non-overlapping waves gives a bootstrap p of 0.080 (Online Resource 1, Fig. S1). The coefficients the report itself publishes in successive editions show the same kind of drift (Table 1), although successive editions share most of their estimation sample, so their differences understate how much the relationship changes across periods.

Further Checks

Details of the checks below are in Online Resource 1, Section S7. Sampling error cannot explain the within-country results: using the published confidence intervals, it accounts for 3.7 per cent of the variance of change in life evaluation between 2019 and 2025 and at most 8.4 per cent in either shorter window.[5] Social support, freedom and perceptions of corruption are measured on the same respondents as life evaluation, so shared sampling error could inflate their within-country slopes. Under generous assumptions the bias is at most about 0.40 ladder points per unit of social support, a tenth of its slope; for freedom it is 0.18, and correcting for it would widen the gap between the within-country and pooled slopes.

Changes in survey mode do not appear to drive the results. Many countries usually surveyed face to face were surveyed by telephone in 2020 and 2021, but the long window compares averages of 2017 to 2019 with averages of 2023 to 2025, and its slopes have the same signs as the three-wave estimates. Slopes differ between high-income and other countries over the long window (asymptotic $p = 0.007$): the point estimates for income (1.45 against 0.35) and freedom (0.78 against 0.29) are larger in high-income countries, where telephone interviewing was used throughout, and the slope for social support is smaller. A single slope describes neither group well.

[5]The sampling variance of a change is the sum of the squared standard errors at the two endpoints, which are independent because the windows share no survey year.

Regression toward the mean does not account for the slopes. Countries that started higher fell more over the long window (coefficient on the 2019 level −0.10, standard error 0.026), but controlling for the starting level leaves the income, social support and freedom slopes nearly unchanged. Timing matters: changes in the factors between 2019 and 2022 predict changes in life evaluation between 2022 and 2025, so the within-country relationship depends on the horizon. With the report's own income series, the within-country income slope between 2022 and 2025 is 0.03 (standard error 0.22), against 1.69 (0.40) with the World Bank series. Finally, the two questions rank countries differently: across the 141 countries observed in 2019 and 2025, the rank correlation between the level of life evaluation in 2025 and its change since 2019 is 0.20, and 21 of the 36 countries in the top quarter by level are below the median on change (Online Resource 1, Section S8).

Discussion

What the Conditions Add

The between-within distinction is old, and the Easterlin paradox has made it familiar for income. The framework proposed here adds two things. It separates the substantive conditions, structural equivalence and temporal invariance, from a measurement condition that is logically prior to both, and it shows that the measurement condition can fail in the most visible well-being instrument in a way that produces a substantive-looking result. Had the probe stopped at the published series, it would have concluded that income and health carry no information about change in national well-being. The conclusion would have been wrong for income. In the published series, the income and health contributions record revisions of their sources at least as much as change in the factors.

The substantive conditions also matter. The probe finds no general rule that within-country relationships are weaker than between-country ones. Social support transfers, income appears to transfer or more than transfer over a six-year window, and freedom does not. The

factors appear to differ in how much of their cross-country association reflects stable features of countries, so a factor that orders countries well may move little with life evaluation within them. Which pattern holds for which factor is an empirical matter, and it can change with the period studied, as the difference between the two windows shows.

The income result bears on the Easterlin debate only in a limited way. The probe covers six years that include a global recession and its recovery, so its within-country income slope is a short-run slope. A short-run slope above the cross-country slope is what Easterlin's own account expects, since he separates fluctuations, which move national happiness, from trends, which on his reading do not (Easterlin & O'Connor, 2022). The probe says nothing about trends. It does show that a test of the paradox built from the report's published series would be biased toward finding it, because revisions to the income series add variation in measured income growth that is unrelated to life evaluation. Long-run tests that splice income series across releases face the same problem, as the growth literature found with successive versions of the Penn World Table (Johnson et al., 2013).

Implications for Reporting and for Progress Dashboards

Three changes would let the decomposition support a dynamic reading. The first is measurement continuity: whenever a country-year series is published, every year should be recomputed with the current edition's coefficients, reference values and data, as official statistics do with back-cast series. The probe shows that this can be approximated from the published files for the four survey-based factors; the report's producers can do it exactly for all six. The second is labelling. The decomposition should be described where it is published as an account of differences between countries, with a statement that its coefficients are not estimated from change within countries. The third is a companion decomposition. The report already estimates a specification with country fixed effects. Publishing contributions to change based on those coefficients, with their uncertainty, alongside the existing bars would

let readers see how much of a country's movement the factors account for and which factors matter for it. A companion table ranking countries by change over a stated window, with confidence intervals, would answer the dynamic question about the outcome itself.

These proposals apply beyond this report, and most directly to progress dashboards. A dashboard that places subjective well-being at its head and reports the conditions that support it below (Burger et al., 2026) invites exactly the reading examined here: when the headline moves, users will look to the conditions for the reason. The three conditions translate into design requirements for such a dashboard. The link between each condition and the headline should be estimated, or at least checked, from change within the unit being tracked, not only from comparisons across units. The link should be re-estimated, and its stability reported, as the series grows. And every year of a published series should rest on the same model and the same data vintage, with revisions applied backwards rather than spliced forwards. Measurement guidance is directed at the headline measure (OECD, 2025); the standard of a consistent series should apply equally to any attribution published with it.

Implications for Research on Change

If the six factors describe differences between countries better than change within them, research aimed at explaining change in national well-being should not start from the cross-sectional model. It should start from within-country designs that allow relationships to vary by horizon and by group of countries, and it should treat the residual term of the decomposition as an object of study rather than a remainder. Work explaining why life evaluations in Latin America are higher than standard predictors imply (Rojas, 2024) shows that the residual can carry substantive content in levels; the probe suggests it carries much of the content in changes as well. Panel evidence that life evaluation responds to events and crises in ways stable characteristics do not anticipate (Kapteyn et al., 2025; Kriechel et al., 2025) points in the same direction. Such research also has to attend to data vintages. Any

analysis of change that uses the report's published contributions inherits the splice documented here.

Limitations

The probe has clear limits. The within-country evidence covers six years and three non-overlapping waves, so it says little about long-run relationships and has limited power. On the wild cluster bootstrap the joint test does not reject the pooled model; the evidence against structural equivalence rests on freedom, and the evidence against temporal invariance rests on the comparison of the two windows. The analysis is at the country level. It can detect heterogeneity in slopes only between broad groups of countries, and it cannot test whether each country has its own slope, which would require individual-level data. The four survey-based factors are rebuilt from published contributions and inherit any revisions to the survey data between editions, and generosity is itself constructed in each edition as a residual from a regression on income, which links it to the income vintage. Life expectancy from the World Bank is a substitute for, not a replica of, the healthy life expectancy series the report uses. The estimates are associations. They establish that the published attribution cannot be read as an account of change and indicate where the substantive conditions fail, but they do not identify the effects of the factors on national well-being.

Conclusion

The World Happiness Report's decomposition answers a comparative question well. Read as an account of change, it fails a condition that is prior to any substantive argument: its published country-year series is assembled from seven editions with different coefficients, reference values and data sources, so year-to-year differences in its contributions mix change with revision. Once the series is rebuilt on a single model and consistent data, the within-country slopes for social support and income are at least as large as their cross-country slopes, although the income estimate is less precise, the slope for freedom is much smaller,

and the within-country relationships shift between the first and second halves of the period. The general lesson is that an attribution estimated from differences between units can be read as an account of change within a unit only if three conditions hold: equivalent relationships, stable relationships and a continuous measurement series. The first two are empirical and partly testable. The third can be checked from the published files and, where it fails, has to be repaired before anything else can be learned. For the World Happiness Report, the producers of the report already hold the data and the estimates needed for that repair.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica, 88*(1), 265–296. https://doi.org/10.3982/ECTA12675

Bartolini, S., & Sarracino, F. (2014). Happy for how long? How social capital and economic growth relate to happiness over time. *Ecological Economics, 108*, 242–256. https://doi.org/10.1016/j.ecolecon.2014.10.004

Bell, A., & Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods, 3*(1), 133–153. https://doi.org/10.1017/psrm.2014.7

Bjørnskov, C. (2010). How comparable are the Gallup World Poll life satisfaction data? *Journal of Happiness Studies, 11*(1), 41–60. https://doi.org/10.1007/s10902-008-9121-6

Bond, T. N., & Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy, 127*(4), 1629–1640. https://doi.org/10.1086/701679

Bromo, F., & Pacek, A. C. (2026). Autocratization and subjective well-being: Is there a macro-level effect? *Applied Research in Quality of Life, 21*(6), 4043–4079. https://doi.org/10.1007/s11482-026-10660-0

621 Brzezinski, M. (2025). Wealth and life satisfaction across Europe: Cross-sectional evidence
622 on satiation and cross-country heterogeneity. *Social Indicators Research, 179*(2), 677–
623 688. https://doi.org/10.1007/s11205-025-03638-2
624 Burger, M., Courchesne, S., Greyling, T., O'Connor, K., Rossouw, S., Sarracino, F., &
625 Veenhoven, R. (2026). Using subjective well-being as a headline indicator in dashboards
626 to track human progress. *Applied Research in Quality of Life, 21*(4), 1735–1761.
627 https://doi.org/10.1007/s11482-026-10576-9
628 Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for
629 inference with clustered errors. *Review of Economics and Statistics, 90*(3), 414–427.
630 https://doi.org/10.1162/rest.90.3.414
631 Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference.
632 *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317
633 Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative income, happiness, and utility: An
634 explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature,*
635 *46*(1), 95–144. https://doi.org/10.1257/jel.46.1.95
636 Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of*
637 *Econometrics, 105*(1), 111–130. https://doi.org/10.1016/S0304-4076(01)00072-0
638 Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person
639 effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619.
640 https://doi.org/10.1146/annurev.psych.093008.100356
641 Deaton, A., & Aten, B. (2017). Trying to understand the PPPs in ICP 2011: Why are the
642 results so different? *American Economic Journal: Macroeconomics, 9*(1), 243–264.
643 https://doi.org/10.1257/mac.20150153
644 De Neve, J.-E., Ward, G., De Keulenaer, F., Van Landeghem, B., Kavetsos, G., & Norton, M.
645 I. (2018). The asymmetric experience of positive and negative economic growth: Global

evidence using subjective well-being data. *Review of Economics and Statistics, 100*(2), 362–375. https://doi.org/10.1162/REST_a_00697

Di Tella, R., MacCulloch, R. J., & Oswald, A. J. (2003). The macroeconomics of happiness. *Review of Economics and Statistics, 85*(4), 809–827. https://doi.org/10.1162/003465303772815745

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David & M. W. Reder (Eds.), *Nations and households in economic growth* (pp. 89–125). Academic Press. https://doi.org/10.1016/B978-0-12-205050-3.50008-7

Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O., & Zweig, J. S. (2010). The happiness–income paradox revisited. *Proceedings of the National Academy of Sciences, 107*(52), 22463–22468. https://doi.org/10.1073/pnas.1015962107

Easterlin, R. A., & O'Connor, K. J. (2022). The Easterlin paradox. In K. F. Zimmermann (Ed.), *Handbook of labor, human resources and population economics* (pp. 1–25). Springer. https://doi.org/10.1007/978-3-319-57365-6_184-2

Exton, C., Smith, C., & Vandendriessche, D. (2015). *Comparing happiness across the world: Does culture matter?* (OECD Statistics Working Papers No. 2015/04). OECD Publishing. https://doi.org/10.1787/5jrqppzd9bs2-en

Frijters, P., Clark, A. E., Krekel, C., & Layard, R. (2020). A happy choice: Wellbeing as the goal of government. *Behavioural Public Policy, 4*(2), 126–165. https://doi.org/10.1017/bpp.2019.39

Graafland, J. (2023). On rule of law, civic virtues, trust, and happiness. *Applied Research in Quality of Life, 18*(4), 1799–1824. https://doi.org/10.1007/s11482-023-10163-2

669 Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework
670 of composite indices: A review of the issues of weighting, aggregation, and robustness.
671 *Social Indicators Research, 141*(1), 61–94. https://doi.org/10.1007/s11205-017-1832-9
672 Greyling, T., & Rossouw, S. (2025). Development and validation of a real-time happiness
673 index using Google Trends™. *Journal of Happiness Studies, 26*(3), Article 39.
674 https://doi.org/10.1007/s10902-025-00881-9
675 Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of*
676 *Econometrics, 31*(1), 93–118. https://doi.org/10.1016/0304-4076(86)90058-8
677 Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46*(6), 1251–1271.
678 https://doi.org/10.2307/1913827
679 Helliwell, J. F., Huang, H., & Wang, S. (2014). Social capital and well-being in times of
680 crisis. *Journal of Happiness Studies, 15*(1), 145–162. https://doi.org/10.1007/s10902-
681 013-9441-z
682 Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.).
683 (2026). *World happiness report 2026*. Wellbeing Research Centre, University of Oxford.
684 https://www.worldhappiness.report/ed/2026/
685 Hill, T. G. (2026). Patterns of national flourishing: Multidimensional clusters of well-being
686 and psychosocial resources across 22 countries. *Applied Research in Quality of Life,*
687 *21*(5), 2875–2894. https://doi.org/10.1007/s11482-026-10610-w
688 Hirata, J., & Groß, P. (2025). Relative effects on life satisfaction revisited: Social comparison
689 is only half the story. *Journal of Happiness Studies, 26*(3), Article 25.
690 https://doi.org/10.1007/s10902-024-00848-2
691 Høyland, B., Moene, K., & Willumsen, F. (2012). The tyranny of international index
692 rankings. *Journal of Development Economics, 97*(1), 1–14.
693 https://doi.org/10.1016/j.jdeveco.2011.01.007

Inglehart, R., Foa, R., Peterson, C., & Welzel, C. (2008). Development, freedom, and rising happiness: A global perspective (1981–2007). *Perspectives on Psychological Science, 3*(4), 264–285. https://doi.org/10.1111/j.1745-6924.2008.00078.x

Johnson, S., Larson, W., Papageorgiou, C., & Subramanian, A. (2013). Is newer better? Penn World Table revisions and their impact on growth estimates. *Journal of Monetary Economics, 60*(2), 255–274. https://doi.org/10.1016/j.jmoneco.2012.10.022

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences, 119*(42), Article e2210412119. https://doi.org/10.1073/pnas.2210412119

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Kapteyn, A., Perez-Arce, F., Sandlin, E. W., & Stone, A. A. (2025). Life satisfaction responds to major life events; affect not so much. *Social Indicators Research, 180*(3), 1747–1763. https://doi.org/10.1007/s11205-025-03724-5

Kriechel, L., Bujard, M., & Hudde, A. (2025). The rollercoaster of subjective well-being in times of multiple crises: Evidence of five waves of bi-annual panel data of the FReDA survey. *Social Indicators Research, 180*(3), 1353–1385. https://doi.org/10.1007/s11205-025-03707-6

Kwarciński, T., Ulman, P., & Wdowin, J. (2024). Measuring subjective well-being capability: A multi-country empirical analysis in Europe. *Applied Research in Quality of Life, 19*(5), 2555–2593. https://doi.org/10.1007/s11482-024-10334-9

Li, R., & Hu, Y. (2026). Identifying key predictors of subjective well-being across 92 countries: A machine learning approach. *Journal of Happiness Studies, 27*(1), Article 16. https://doi.org/10.1007/s10902-025-01004-0

718 Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). Subjective well-being and
719 adaptation to life events: A meta-analysis. *Journal of Personality and Social Psychology,*
720 *102*(3), 592–615. https://doi.org/10.1037/a0025948
721 Mader, S., & Franzen, A. (2025). Friends make us happy: Evidence from three European
722 panel studies. *Journal of Happiness Studies, 26*(4), Article 55.
723 https://doi.org/10.1007/s10902-025-00888-2
724 Messick, S. (1995). Validity of psychological assessment: Validation of inferences from
725 persons' responses and performances as scientific inquiry into score meaning. *American*
726 *Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741
727 Mikucka, M., Sarracino, F., & Dubrow, J. K. (2017). When does economic growth improve
728 life satisfaction? Multilevel analysis of the roles of social trust and income inequality in
729 46 countries, 1981–2012. *World Development, 93*, 447–459.
730 https://doi.org/10.1016/j.worlddev.2017.01.002
731 Mueller, G. P. (2023). The recent dangers for European happiness: Is homeostatic resilience
732 sufficient? *Applied Research in Quality of Life, 18*(4), 2089–2105.
733 https://doi.org/10.1007/s11482-023-10178-9
734 Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica,*
735 *46*(1), 69–85. https://doi.org/10.2307/1913646
736 Nilsson, A. H., Eichstaedt, J. C., Lomas, T., Schwartz, A., & Kjell, O. (2024). The Cantril
737 Ladder elicits thoughts about power and wealth. *Scientific Reports, 14*, Article 2642.
738 https://doi.org/10.1038/s41598-024-52939-y
739 Ochoa Pacheco, P., Coello-Montecel, D., & Tello, M. (2026). Less income and work, less
740 well-being? Evidence on the impact of the Covid-19 pandemic on well-being from the
741 Gallup World Poll. *Applied Research in Quality of Life, 21*(2), 545–574.
742 https://doi.org/10.1007/s11482-025-10534-x

OECD. (2025). *OECD guidelines on measuring subjective well-being (2025 update)*. OECD Publishing. https://doi.org/10.1787/9203632a-en

Oishi, S., & Kesebir, S. (2015). Income inequality explains why economic growth does not always translate to an increase in happiness. *Psychological Science, 26*(10), 1630–1638. https://doi.org/10.1177/0956797615596713

Padgett, R. N., Felton, C., Case, B., Lomas, T., Johnson, B., & VanderWeele, T. (2026). Life satisfaction is more strongly correlated with flourishing than Cantril's ladder. *Applied Research in Quality of Life, 21*(5), 3337–3348. https://doi.org/10.1007/s11482-026-10623-5

Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society Series A: Statistics in Society, 176*(3), 609–634. https://doi.org/10.1111/j.1467-985X.2012.01059.x

Ponomareva, N., & Katayama, H. (2010). Does the version of the Penn World Tables matter? An analysis of the relationship between growth and volatility. *Canadian Journal of Economics, 43*(1), 152–179. https://doi.org/10.1111/j.1540-5982.2009.01567.x

Prati, G. (2022). The association between subjective well-being and regime type across 78 countries: The moderating role of political trust. *Applied Research in Quality of Life, 17*(6), 3393–3413. https://doi.org/10.1007/s11482-022-10070-y

Ravallion, M. (2012). Mashup indices of development. *The World Bank Research Observer, 27*(1), 1–32. https://doi.org/10.1093/wbro/lkr009

Röck, S., Kleinheinz, L., & Walde, J. (2026). Assessing measurement invariance of subjective well-being across European countries. *Journal of Happiness Studies, 27*(2), Article 37. https://doi.org/10.1007/s10902-026-01011-9

766 Rojas, M. (2024). The joint enjoyment of life: Explaining high happiness in Latin America.
767 *Journal of Happiness Studies, 25*(7), Article 100. https://doi.org/10.1007/s10902-024-
768 00817-9
769 Sacks, D. W., Stevenson, B., & Wolfers, J. (2012). The new stylized facts about income and
770 subjective well-being. *Emotion, 12*(6), 1181–1187. https://doi.org/10.1037/a0029873
771 Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis
772 techniques as tools for the quality assessment of composite indicators. *Journal of the*
773 *Royal Statistical Society Series A: Statistics in Society, 168*(2), 307–323.
774 https://doi.org/10.1111/j.1467-985X.2005.00350.x
775 Schilpzand, A., & Graafland, J. (2026). Life satisfaction in an unequal world: On the
776 moderating role of GDP per capita and national culture. *Applied Research in Quality of*
777 *Life, 21*(5), 3167–3200. https://doi.org/10.1007/s11482-026-10613-7
778 Smith, M. D., & Wesselbaum, D. (2023). Well-being and income across space and time:
779 Evidence from one million households. *Journal of Happiness Studies, 24*(5), 1813–1840.
780 https://doi.org/10.1007/s10902-023-00660-4
781 Stevenson, B., & Wolfers, J. (2008). Economic growth and subjective well-being:
782 Reassessing the Easterlin paradox. *Brookings Papers on Economic Activity, 2008*(1), 1–
783 87. https://doi.org/10.1353/eca.0.0001
784 Tofallis, C. (2020). Which formula for national happiness? *Socio-Economic Planning*
785 *Sciences, 70*, Article 100688. https://doi.org/10.1016/j.seps.2019.02.003
786 Veenhoven, R. (2026). How to take stock of research findings on happiness in regions using
787 the World Database of Happiness. *Applied Research in Quality of Life, 21*(1), 9–33.
788 https://doi.org/10.1007/s11482-024-10402-0
789 World Bank. (2026). *World development indicators* [Data set; series NY.GDP.PCAP.PP.KD
790 and SP.DYN.LE00.IN, last updated 8 October 2026]. https://api.worldbank.org/v2/

World Happiness Report. (2026). *Data for Figure 2.1* [Data set; editions 2020 to 2026]. https://www.worldhappiness.report/data-sharing/

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf
- ESM2.xlsx
- ESM3.zip