WELL-BEING IN A SUSTAINABLE ECONOMY REVISITED

WISER Database

WP 2, Task 2.3, Deliverable 2.3

February 2025

WISER: Well-being in a Sustainable Economy Revisited

WISER – 101094546

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101094546.

Project Information

Project acronym:	WISER
Full title of project:	Well-being in a Sustainable Economy Revisited
Call identifier:	HORIZON-CL2-2022-TRANSFORMATIONS-01
Type of action:	RIA
Start date:	1 October 2023
End date:	30 September 2026
Grant agreement no:	101094546

Deliverable 2.3 – WISER Database

WP 2:	WISER Database
Due Date:	28-2-2025
Submission Date:	28-2-2025
Responsible Partner:	OUNL
Version:	1.0
Status:	Final
Author(s)	
Deliverable Type:	R
Dissemination Level:	PU

Statement of originality

This report contains original unpublished work except were indicated otherwise. The work of others and published material has been indicated through citation, quotation or both.

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Agency. Neither the European Union nor the granting authority can be held responsible for them.

Version History

Version	Date	Author	Partner	Description
0.0A	26-02-2025	Martijn Burger, Spyridon Stavropoulos, Francesco Sarracino, Sarah Courchesne	OUNL, UPATRAS, STATEC	Draft for Part A
0.0B	01-11-2024	Talita Greyling, Stephanié Rossouw	UJ, AUoT	Draft for Part B
0.0C	01-12-2024	Spyridon Stavropoulos, Martijn Burger	UPATRAS, OUNL	Draft for Part C
1.0	28-02-2025	Sarah Courchesne, Martijn Burger	OUNL	Final Version

Abbreviations and Acronyms

Abbreviation	Fully written
AI	Artificial Intelligence
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
API	Application Programming Interface
CGSS	Chinese General Social Survey
DV	Dependent Variable
EB	Eurobarometer
BHP	British Household Panel
BRFSS	Behavioural Risk Factor Surveillance System
EARS	Response with Social Listening
EQLS	European Quality of Life Survey
ESS	European Social Survey
GDP	Gross Domestic Product
GNH	Gross National Happiness
GSS	General Social Survey
GSSR	The Graduate School for Social Research
GWP	Gallup World Poll
HH	Household
HILDA	Household, Income and Labour Dynamics in Australia
IFis PAN	Institute of Philosophy and Sociology of the Polish Academy of Sciences
LITS	Life in Transition Survey
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
NA	Negative Affect

NLP	Natural Language Processing
NRC	National Research Council
NUTS	Territorial Units for Statistics
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinal Least Squares
ONS	Office for National Statistics
PA	Positive Affect
PANAS	Positive and Negative Affect Schedule
RMSE	Root Mean Square Error
RSV	Relative Search Volume
SD	Standard Deviation
SOEP	German Socio-economic Panel
SPANE	Scale of Positive and Negative Experience
SWBI	Subjective Well-being Index
SWLS	Satisfaction with Life Scale
UK	United Kingdom
VADER	Valence Aware Dictionary and Sentiment Reasoner
WHO	World Health Organisation
WISER	Well-being in a Sustainable Economy Revisited
WVS	World Values Survey

Table of Contents

List of Figures

List of Tables

Summary

The data package is accompanied by two papers and a pre-study (Appendix A). The data package itself will be made publicly available with the dashboard.

In part 1, we present a new dataset on subjective well-being in European regions. Regional SWB data is thin on the ground. Over the past years, cross-national SWB data in Europe has been collected in several large surveys such as the Gallup World Poll, Eurobarometer, European Social Survey, and European Values Survey. Despite increasing interest in subnational SWB analyses, annual data collections are typically only representative at the national level and not at the regional level. A solution here could be to pool and harmonize all available SWB data from different surveys, which is for example done in the World Database of Happiness and the survey data recycling project by IFiS PAN and GSSR PAN.

In the first paper, we present a new cross-regional and panel data on subjective well-being in European regions. We first present cross-regional statistics for the period 2005-2024 based on several subjective well-being measures and relying on long-term cross-sectional data collections such as the Eurobarometer, European Social Survey, European Values Survey, and Gallup World Poll. Subsequently, we construct a panel database on annual evaluative well-being in European regions by combining data from different cross-sectional surveys and using a dynamic Bayesian latent variable model (Claassen, 2019; 2022), which allows us to estimate regional evaluative well-being across multiple regions and time points. This model is well-suited for aggregating survey responses over numerous regions and years, particularly when existing survey data are fragmented across various survey questions and contain significant gaps in each regional time series. Specifically, the dynamic Bayesian latent variable model corrects for the bias induced by the usage of particular items in particular regions. In the pre-study (Appendix A), we examined to what extent different question wording generate different outcomes.

In part 2, we present a new way of measuring subjective well-being. Traditionally, surveys have been the main method for measuring happiness, but they face challenges such as "survey fatigue", high costs, time delays, and the fluctuating nature of happiness. Addressing these challenges of survey data, Big Data from sources like Google Trends™ and social media is now being used to complement surveys and provide policymakers with more timely insights into well-being.

In recent years, Google Trends™ data has been leveraged to discern trends in mental health, including anxiety and loneliness, and construct robust predictors of subjective well-being composite categories. We aim to construct the first comprehensive, near real-time measure of population-level happiness using information-seeking query data extracted continuously using Google Trends™. We use a basket of English-language emotion words suggested to capture positive and negative affect and apply machine learning algorithms—XGBoost and ElasticNet—to identify the most important words and their weight in estimating happiness.

We demonstrate our methodology using data from the United Kingdom and test its cross-country applicability in the Netherlands by translating the emotion words into Dutch. Lastly, we improve the fit for the Netherlands by incorporating country-specific emotion words. Evaluating the accuracy of our estimated happiness in countries against survey data, we find a very good fit with very low error metrics. Adding country-specific words improves the fit statistics. Our suggested innovative methodology demonstrates that emotion words extracted from Google Trends™ can accurately estimate a country's level of subjective well-being. The method can be easily implemented in other countries and regions.

Part A- A New Database on Subjective Well-Being for European Regions

1 Introduction

In 1971, King Jigme Singye Wangchuck of Bhutan declared that 'Gross National Happiness' (GNH) is more significant than 'Gross National Product' (GNP). Over the following decades, the notion that GNP is an inadequate measure of a nation's success has gained traction beyond Bhutan. In 2012, the United Nations General Assembly passed a resolution urging governments to prioritize the subjective well-being of their citizens. Similarly, in England, France, and Germany, the administrations of Cameron, Sarkozy, and Merkel announced that enhancing happiness would become a key governmental objective. However, the focus on happiness as a policy issue is not limited to national governments. Many regions have also begun to develop their own 'Gross Regional Happiness' frameworks, shifting their focus toward quality-of-life indicators and subjective well-being, thereby moving 'Beyond GDP'.

Why is there growing interest in placing subjective well-being at the forefront of policy agendas? First, there is increasing public recognition of the importance of happiness, which has become a key aspiration for citizens in the 21st century. According to the OECD Better Life Index, life satisfaction ranks just behind health as the second most important factor for a better life, surpassing education, income, and civic engagement. Second, local governments are increasingly competing to attract highly educated workers and large corporations. Since subjective well-being reflects the subjective assessment of quality of life, regions with higher levels of well-being are more likely to be attractive places to live and work (e.g., Lenzi & Perucca, 2018; Burger et al., 2020). Third, higher subjective well-being (SWB) is associated with better physical and mental health and increased productivity, which can lead to lower healthcare costs and improved economic performance (e.g., Veenhoven, 2008; De Neve et al., 2013; Fang et al., 2024). Fourth, happier citizens are more likely to support incumbent governments in local elections, providing a political incentive for leaders to prioritize well-being (e.g., Liberini et al., 2017; Ward, 2020; Burger & Eiselt, 2023).

However, if regional governments are tasked with enhancing the happiness of their citizens, it is essential to develop tools to measure their success in achieving this goal. Metrics such as Gross Regional Happiness are not only vital for assessing public policy effectiveness but also serve as democratic instruments for evaluating governance. In this context, it is crucial not only to measure current levels of citizen SWB but also to determine whether local governments can effectively raise SWB and identify the strategies that make this possible.

Yet, regional SWB data is thin on the ground. Over the past years, cross-national SWB data in Europe has been collected in several large surveys such as the Gallup World Poll, Eurobarometer, European Social Survey, and European Values Survey. Despite increasing interest in subnational SWB analyses (e.g., Berry & Okulicz-Kozaryn, 2011; Lenzi & Perucca, 2018; Burger et al., 2020), annual data collections are typically only representative at the national level and not at the regional level. A solution here could be to pool and harmonize all available SWB data from different surveys, which is for example done in the World Database of Happiness (Veenhoven, 2024; 2025) and the survey data recycling project by IFiS PAN and GSSR PAN (Slomzynski et al., 2022).

In this article, we present a new cross-regional and panel data on subjective well-being in European regions. We first present cross-regional statistics for the period 2005-2024 presenting several subjective well-being measures, relying on long-term cross-sectional data collections such as the Eurobarometer, European Social Survey, European Values Survey, and Gallup World Poll. Subsequently, we construct a panel database on annual evaluative well-being in European regions by combining data from different cross-sectional surveys and using a dynamic Bayesian latent variable model (Claassen, 2019; 2022), which allows us to estimate regional evaluative well-being across multiple regions and time points. This model is well-suited for aggregating survey responses over numerous regions and years, particularly when existing survey data are fragmented across various survey questions and contain significant gaps in each regional time series. Specifically, the dynamic Bayesian latent variable model corrects for the bias induced by the usage of particular items in particular regions.

The remainder of this paper is organized as follows. Section 2 discusses the measurement of subjective well-being and the available measures and methods to construct the long-term panel. Section 3 presents the data. Section 4 concludes.

2 Data and Panel Construction

2.1 Measuring Subjective Well-being

Subjective well-being can be defined as '*the degree to which an individual judges the overall quality of his/her own life-as-a-whole favorably*' (Veenhoven, 1984)'. An outsider often cannot assess how someone internally experiences their life, and it is currently impossible to measure subjective well-being through bodily processes. While we can measure stress through hormone levels and blood pressure, we cannot measure subjective well-being. Since all mental experiences ultimately have a physiological substrate, it should in the future be possible to measure subjective well-being through brain scans (see e.g. Van 't Ent et al., 2017). However, until then we will have to rely on self-reports. This is not necessarily a bad thing, because happiness is, by definition, something that people experience in their minds, and thoughts are best measured by asking individuals about them.

The OECD SWB into evaluative, affective, and eudaimonic measures. Evaluative measures of subjective well-being involve people's overall assessments of their lives or specific aspects of it, commonly captured by asking respondents to rate their life satisfaction. Contentment is generally included in this category. Affective measures, on the other hand, focus on people's feelings and emotions, often assessed over a specific time period (such as “yesterday”). Eudaimonia refers to psychological flourishing and is measured by how much individuals feel their lives have purpose or meaning, including elements like autonomy, competence, and self-actualization (Mahoney, 2023). While eudaimonic measures are often used in positive psychology, they are less common in economics and sociology. This is partly because many well-being researchers view these aspects as conditions for a good life rather than outcomes of it (Veenhoven, 2017). Martela (2025) suggests that the different elements of eudaimonic well-being, which reflects human functioning, are important antecedents for both evaluative and hedonic aspects of well-being.

2.2 Cross-regional Data Collection

In the database of subjective well-being, we present evaluative, affective, and eudaimonic measures for European regions. These measures are derived from several repeated cross-section surveys. However, given that most of these data collections are only representative at the national level, it is not possible to present annual measures and information from different survey rounds need to be combined to get accurate measures of SWB at the regional level.

In this database, we present subjective well-being for a combination of NUTS-1 and NUTS-2 regions. Not all member states have similar shaped NUTS regions at all levels. Smaller countries in Europe, such as Cyprus and Luxembourg have no subdivision in NUTS-1 and NUTS-2 areas because their population size of less than 3 million does not impose a NUTS 1 or 2 breakdown. All countries are given two-letter abbreviation; a NUTS region is designated by that two-letter abbreviation followed by a numerical designator. For example, in Italy ITC is Northwest Italy and ITC1 is Piemonte. In the estimation of SWB scores, we omit regions with less than 100 observations because that may lead to imprecise estimates of the SWB score. We harmonized the regional classification to the NUTS-2024 division.

We obtain our cross-sectional regional SWB data for the period 2005-2024 from three databases: Eurobarometer (1,061,005 respondents), Gallup World Poll (530,966 respondents), and the European Social Survey (387,884 respondents). Other surveys such as the World Values Survey or European Quality-of-Life Survey contained not enough respondents across different waves to provide accurate regional aggregate scores. However, we will use these surveys in the modelling of cross-regional time-series on SWB. From the Eurobarometer (EB), Gallup World Poll (GWP) and European Social Survey (ESS) the following regional SWB measures are constructed:

Table 1. Cross-regional SWB measures

Survey	Type of measure	Measurement
EB	Evaluative well-being	On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead? 1 (very satisfied) – 4 (not at all satisfied). Rescaled to a 0-10 measure.
GWP	Evaluative well-being	Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?
	Affective well-being	Positive affect Did you feel well-rested yesterday? (Yes/No) Were you treated with respect all day yesterday? (Yes/No) Did you smile or laugh a lot yesterday? (Yes/No) Did you learn or do something interesting yesterday? (Yes/No) Did you experience enjoyment yesterday? (Yes/No) Negative affect Did you experience the following feelings during a lot of the day yesterday? How about physical pain? (Yes/No) How about worry? (Yes/No) How about sadness? (Yes/No) How about stress? (Yes/No) How about anger? (Yes/No)

		Affect Balance = Positive Affect – Negative Affect. Score: Yes=1, No =0, Rescaled to 0 to 10 scale.
	Eudaimonic well-being (Joshanloo et al., 2018)	Did you learn or do something interesting yesterday? (Yes/No) If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not? (Yes/No) Were you treated with respect all day yesterday (Yes/No) Can people in this country get ahead by working hard, or not? (Yes/No) In this country, are you satisfied or dissatisfied with your freedom to choose what to do with your life? (Satisfied/Dissatisfied) In the past month, have you helped a stranger or someone you didn't know who needed help? (Yes/No) In the past month, have you volunteered your time to an organization? (Yes/No); Score: Yes=1, No =0, Rescaled to 0 to 10 scale.
ESS	Evaluative well-being – life satisfaction	All things considered, how satisfied are you with your life as a whole nowadays? 0 (extremely dissatisfied) - 10 (extremely satisfied)
ESS	Evaluative well-being - happiness	Taking all things together, how happy would you say you are? 0 (extremely unhappy) - 10 (extremely happy)

2.3 Modelling Cross-regional Time Series

There are at present many cross-national survey projects in Europe, regularly asking respondents about their evaluative well-being – unfortunately measurements of affective and eudaimonic well-being are thinner on the ground. However, some regions have not been surveyed at one time or another, and many regions have been surveyed numerous times, sometimes by several of these survey projects. Yet, no single survey provides sufficient information to create a panel dataset. However, the fact that there exist several of these survey projects that have been running for decades, this offers the possibility to construct the possibility of constructing a panel of evaluative well-being for European regions. Unfortunately, regional evaluative well-being has fragmented time series and large spatial gaps, and some regions are underrepresented in different ways. In addition, evaluative well-being is measured in different ways using overall happiness, life satisfaction and Cantril ladder measures. In addition, there can be considerable survey effects. As part of preparing the data package, we examined the effects of question wording on outcomes (see Appendix B). Using experimental data from over 7,700 respondents across four experiments, the findings show that the wording of single-item happiness and life satisfaction measures has limited effects. While lower SWB and higher dispersion is occasionally observed in SWB questions with rarely used neutral or negative question tones, more common wording differences do not affect the predictors, means, and dispersion of SWB. The findings imply that while the lack of a uniform SWB measure is not ideal, it does not pose a significant threat to the credibility of findings from the SWB literature. Yet, there are still survey effects and different types of questions may still capture different facets of evaluative well-being scores. While the Cantril ladder particularly captures the cognitive dimensions of evaluative well-being, overall happiness also contains notions of emotional well-being.

This database therefore utilizes a method for estimating smooth panels of aggregate evaluative well-being using all available survey data. As input, we rely on all measurements of evaluative well-being that are available in cross-sectional databases. An overview of the items included can be found in Table 2. As can be observed from this table, most items focus on life satisfaction and overall happiness – the Cantril ladder question can only be found for multiple years in the Gallup World Poll. To model cross-regional temporal variations in subjective well-being, we apply the Bayesian latent variable modelling approach developed by Claassen (2019, 2022). First, evaluative well-being is treated as a latent, unobserved characteristic. This means that the overall levels of evaluative well-being are shaped both by the actual well-being state and by measurement errors stemming from factors such as the survey framework, item phrasing, response scales, and random variation. Second, we adjust for differences in item functioning across regions by incorporating factor loading scores for regional items within the measurement model. Third, to account for variations in respondent samples across different survey projects, we incorporate a specification for sampling error. Lastly, while measurement models help capture the unobserved trait by correcting for error-related variance, data gaps still exist. To address this, we apply temporal smoothing, modelling the latent evaluative well-being level through a local-level dynamic linear model, where the score at time t depends on the score at t-1, with added random noise.

Table 2. Survey-Item-Scale-Year Overview

Survey	Question	Scale	Years
Measure A: Life satisfaction			
EQLS	All things considered, how satisfied would you say you are with your life these days?	1 (very dissatisfied) – 10 (very satisfied)	2008, 2012, 2016
ESS	All things considered, how satisfied are you with your life as a whole nowadays?	0 (extremely dissatisfied) - 10 (extremely satisfied)	2006, 2008, 2010, 2012, 2014, 2016, 2018
EB	On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead?	1 (very satisfied) – 4 (not at all satisfied)	2005-2023
EVS/WVS	All things considered, how satisfied are you with your life as a whole these days?	1 (completely dissatisfied) - 10 (completely satisfied)	2005-2009, 2011-2013, 2017-2022
GWP	All things considered, how satisfied are you with your life as a whole these days?	0 (dissatisfied) -10 (satisfied)	2008-2010
LITS	All things considered, I am satisfied with my life now	1 (completely dissatisfied - 10 (completely satisfied)	2010, 2016
Measure B: Overall happiness			
ESS	Taking all things together, how happy would you say you are?	0 (extremely unhappy) - 10 (extremely happy)	2005-2024

EVS/WVS	Taking all things together, would you say you are:	1 (very happy) - 4 (not at all happy)	2005-2009, 2011-2013, 2017-2022
EQLS	Taking all things together on a scale of 1 to 10, how happy would you say you are?	1 (very unhappy) – 10 (very happy)	2008, 2012, 2016
Measure C: Cantril ladder			
GWP	Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?	0 (worst possible life) – 10 (best possible life)	2005-2024

Note: WVS=World Values Survey; EVS=European Values Survey; ESS=European Social Survey; EB=Eurobarometer; GWP=Gallup World Poll; EQLS=European Quality of Life Survey; LITS= Life in Transition Survey

3 Data Features

Table 3 provides the average regional scores for each of the aggregate measures based on the Eurobarometer, Gallup World Poll and European Social Survey. From the Table, it can be observed that all averages for European regions lie within the range of 6.3-7.2 and all evaluative measures lie within a range of a 6.3-6.7. In addition, the correlation between the different measures is high to very high – not only between the different evaluative wellbeing measures, but also between the evaluative, emotional and eudaimonic measures.

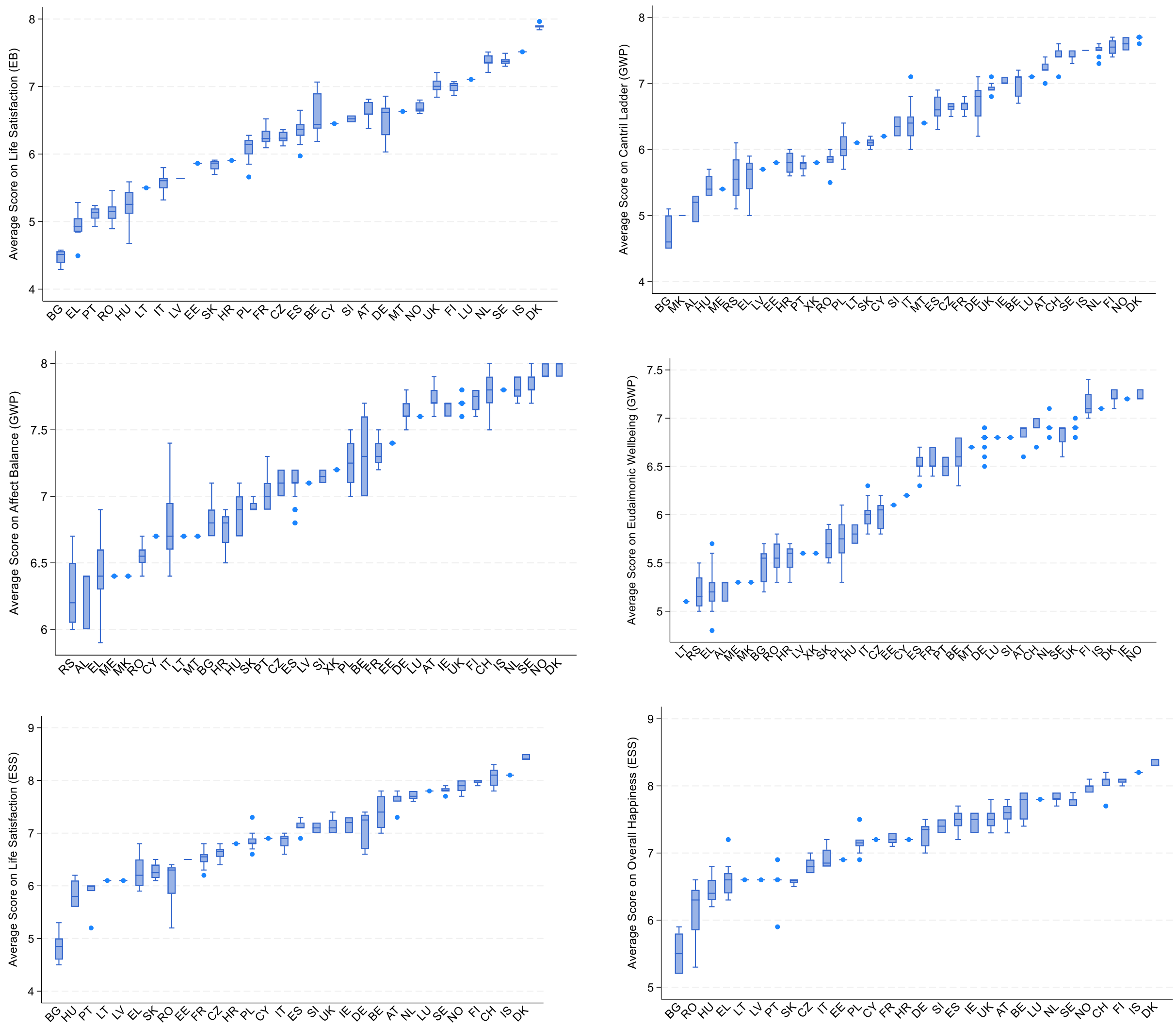
Table 3. Regional Averages of Subjective Well-being 2005-2024 and Correlation Matrix

	Mean score (original)	Number of regions	(1)	(2)	(3)	(4)	(5)	(6)
(1) EB – Life satisfaction	6.3	195	1.00					
(2) GWP-Cantril	6.6	206	0.91	1.00				
(3) GWP-Affect balance	7.2	206	0.85	0.84	1.00			
(4) GWP-Eudaimonic	6.3	206	0.82	0.86	0.86	1.00		
(5) ESS – Life satisfaction	6.7	194	0.90	0.91	0.77	0.76	1.00	
(6) ESS – Happiness	6.6	194	0.90	0.89	0.78	0.79	0.96	1.00

Note: EB=Eurobarometer, GWP=Gallup World Poll, ESS=European Social Survey

Figure 1 show the average scores by country and regional variation within countries. In line with earlier studies, the regions with the highest well-being scores can be found in Northern Europe, while the least happy regions can be found in Eastern and Southern Europe. The regional variation of scores is generally higher in Eastern and Southern Europe than in Northwest Europe.

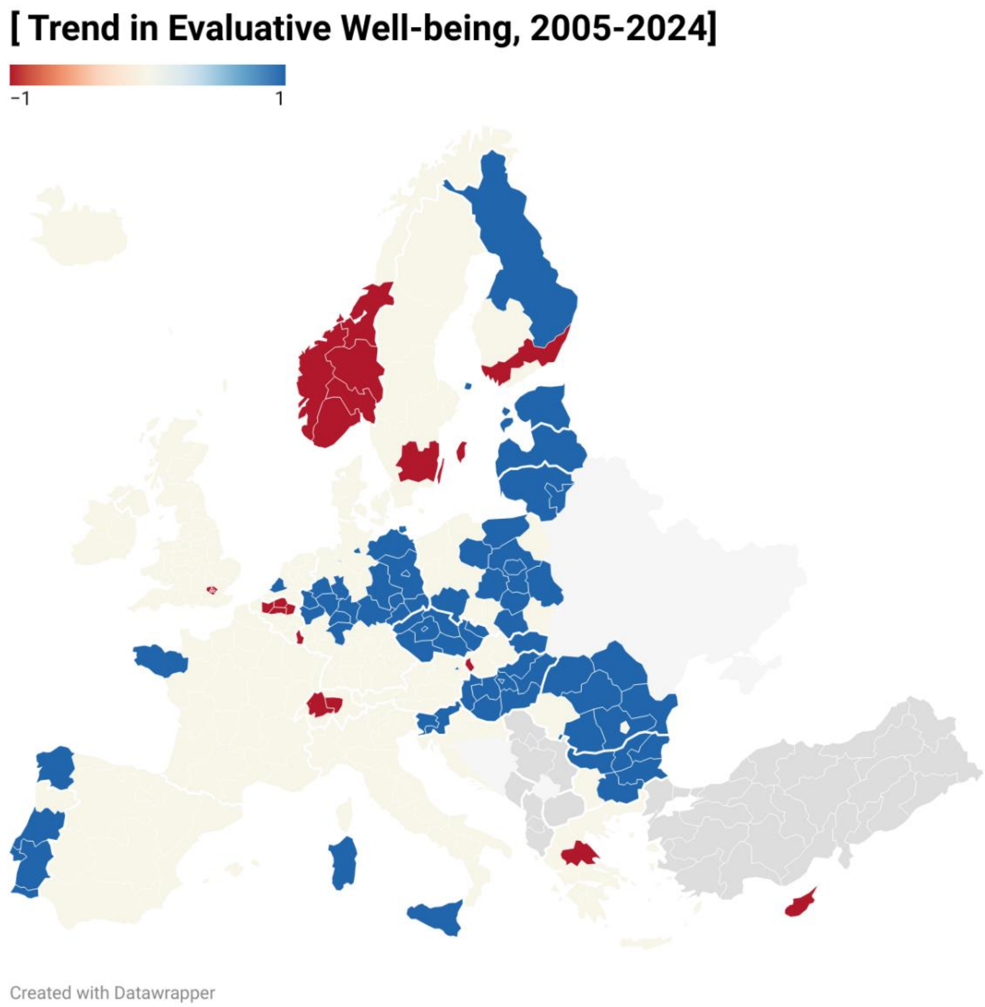
Figure 1. Life Satisfaction in European Regions- Eurobarometer



When combining all evaluative well-being data in the created panel, we observe

Evaluating the trend in evaluative well-being within regions, we observe convergence between Eastern and Western Europe (Figure 2). Where evaluative well-being in many East European regions is increasing, evaluative well-being in most West European regions is stable or decreasing. Future research should examine the drivers behind these changes.

Figure 2: Development of evaluative well-being in European regions



4 Concluding Remarks

In this paper, we have presented a new cross-sectional and panel database on subjective well-being in European regions, based on different databases. This database can help to address open questions in the field, relating to (e.g.) convergence in well-being across European regions, the effects of European funding on subjective well-being, and whether well-being predicts important outcomes such as voting behavior and productivity at the regional level. An overview of other variables in the data package (to be updated in the future) can be found in Appendix B.

References

Berry, B. J., & Okulicz-Kozaryn, A. (2011). An urban-rural happiness gradient. *Urban Geography*, 32(6), 871-883.

Burger, M. J., & Eiselt, S. (2023). Subjective well-being and populist voting in the Netherlands. *Journal of Happiness Studies*, 24(7), 2331-2352.

Burger, M. J., Morrison, P. S., Hendriks, M., & Hoogerbrugge, M. M. (2020). Urban-rural happiness differentials across the world. *World Happiness Report*, 2020, 66-93.

Claassen, C. (2019). Estimating smooth country-year panels of public opinion. Political Analysis, 27(1), 1- 20.

Claassen, C. (2022). Measuring Democratic Mood: Methods. Mimeo.

De Neve, J. E., Diener, E., Tay, L., & Xuereb, C. (2013). The objective benefits of subjective well-being. EP Discussion Papers (CEPDP1236). London School of Economics and Political Science. Centre for Economic Performance, London, UK.

Joshanloo, M. (2018). Optimal human functioning around the world: A new index of eudaimonic well-being in 166 nations. *British Journal of Psychology*, 109(4), 637-655.

Lenzi, C., & Perucca, G. (2018). Are urbanized areas source of life satisfaction? Evidence from EU regions. *Papers in Regional Science*, 97, S105-S122.

Liberini, F., Redoano, M., & Proto, E. (2017). Happy voters. Journal of Public Economics, 146, 41-57.

Martela, F. (2025). Well-Being as Having, Loving, Doing, and Being: An Integrative Organizing Framework for Employee Well-Being. *Journal of Organizational Behavior*, 1-21.

Slomczynski, K. M., Tomescu-Dubrow, I., & Wysmulek, I. (2022). Survey data quality in analyzing harmonized indicators of protest behavior: A survey data recycling approach. *American Behavioral Scientist*, 66(4), 412-433.

Veenhoven, R. (2008). Healthy happiness: Effects of happiness on physical health and the consequences for preventive health care. *Journal of happiness studies*, *9*, 449-469.

Veenhoven, R. (2017). Measures of happiness: Which to choose?. In: *Metrics of subjective well-being: Limits and improvements* (pp. 65-84).

Veenhoven, R. (2024). World database of Happiness. In *Encyclopedia of quality of life and well-being research* (pp. 7861-7865). Cham: Springer International Publishing.

Veenhoven, R. (2025). How to take stock of research findings on happiness in regions using the World Database of Happiness. *Applied Research in Quality of Life*, 1-25.

Ward, G. (2020). Happiness and voting: Evidence from four decades of elections in Europe. *American Journal of Political Science*, *64*(3), 504-518.

Part B- Development and Validation of a Real-time Happiness Index Using Google Trends™

Abstract

It is well-established that a positive relationship exists between happiness and the economic outcomes of a country. Traditionally, surveys have been the main method for measuring happiness, but they face challenges such as "survey fatigue", high costs, time delays, and the fluctuating nature of happiness. Addressing these challenges of survey data, Big Data from sources like Google Trends™ and social media is now being used to complement surveys and provide policymakers with more timely insights into well-being.

In recent years, Google Trends™ data has been leveraged to discern trends in mental health, including anxiety and loneliness, and construct robust predictors of subjective well-being composite categories. We aim to construct the first comprehensive, near real-time measure of population-level happiness using information-seeking query data extracted continuously using Google Trends™. We use a basket of English-language emotion words suggested to capture positive and negative affect and apply machine learning algorithms—XGBoost and ElasticNet—to identify the most important words and their weight in estimating happiness.

We demonstrate our methodology using data from the United Kingdom and test its cross-country applicability in the Netherlands by translating the emotion words into Dutch. Lastly, we improve the fit for the Netherlands by incorporating country-specific emotion words.

Evaluating the accuracy of our estimated happiness in countries against survey data, we find a very good fit with very low error metrics. Adding country-specific words improves the fit statistics. Our suggested innovative methodology demonstrates that emotion words extracted from Google Trends™ can accurately estimate a country's level of happiness.

1 Introduction

Measuring well-being using subjective measures is essential since it is accepted that a positive relationship exists between happiness and the economic outcomes of a country. People's happiness profoundly affects these outcomes, including productivity, labour market performance, and future income (Piekalkiewicz, 2017; Bryson et al., 2016). Increased happiness also positively affects a nation's social and health sectors (Kim et al., 2015), fosters altruistic behaviour and enhances various cognitive and social capabilities (Kasser & Ryan, 1996; Williams & Shiaw, 1999). Happier individuals are healthier, live longer, and generally report higher levels of life satisfaction. They are more likely to avoid high-risk activities and take preventive measures to reduce potential risks.

Traditionally, the primary source for measuring people's happiness has been survey data. However, in a post-pandemic era, people experience 'survey fatigue'. Moreover, conducting surveys is expensive and often results in data that is delayed by up to two years, which may also be affected by non-response bias (Callegaro & Yang, 2018; Rossouw & Greyling, 2020).

To overcome these limitations of survey data, researchers have turned to Big Data to measure and track people's happiness. Measuring people's happiness using Big Data adds an additional benefit since decision-makers are often confronted with short-term horizons and imperfect information. Therefore, they need an immediate source of information regarding a country's mood so that people's needs and concerns guide policies for achieving collective outcomes (Rossouw & Greyling, 2024). Real-time information from Big Data will also allow decision-makers to gauge possible reactions to the proposed legislature to mitigate potentially violent and destructive outcomes (Greyling & Rossouw, 2022). Notably, the work done by Dodd and Danforth (2010), Iacus et al. (2015 and 2022), and Greyling and Rossouw (2019) set the way to harness the power of Big Data. All three studies utilised Twitter data to construct happiness or subjective well-being measures (see section 2.2 for full discussion).

Unfortunately, with Elon Musk's purchase of Twitter (now X), all academic licenses were suspended, and access to Twitter data was stopped, effectively closing the book on academic research. Therefore, researchers focusing on measuring happiness in real time had to resort to other Big Data sources.

Recent research has demonstrated the value of information-seeking query data in forecasting social phenomena. Carammia et al. (2022) utilised a Dynamic Elastic Net (DynENet) model to predict asylum-related migration flows by integrating administrative data with non-traditional sources such as internet searches and geolocated event data. Their work has underscored the broader applicability of internet-derived data for real-time social monitoring.

Therefore, we aim to explore an innovative methodology to accurately estimate happiness levels and their evolution at a country level from information-seeking query data based on a carefully curated selection of English emotion words extracted continuously from Google Trends™. In our proof of concept, we validate our index using the happiness survey measure from the United Kingdom's (UK) Office for National Statistics (ONS) (referred to as True Happiness). Our second aim is to explore whether the same selected basket of English words translated into a different language (Dutch) with the same weights can also successfully estimate happiness in the Netherlands. Here, we validate our equation against the Dutch Time Use data. Our last aim is to follow the same methodology for our initial derived UK happiness equation using the Dutch Time Use happiness measure as the outcome variable for predictions. Here, we use the initial basket of emotion words and add country-specific words to attain a more accurate estimate of happiness in a country.

Previous studies (see section 2.3 for full discussion) leveraging Google Trends™ data measured trends in mental health, including depression, anxiety, and loneliness (Foa et al., 2022; Brodeur et al., 2020; Ford et al., 2018), constructed robust predictors of subjective well-being composite categories in the United States (Algan et al., 2019) and nowcasted national average subjective well-being (Murtin & Salomon-Ermel, 2024). However, none of these studies attempted to estimate county-level happiness measured and updated in almost real-time. Therefore, to our knowledge, our near real-time measure of population-level happiness using information-seeking query data extracted continuously using Google Trends™ in countries is a first of its kind.

To construct our happiness index, we start by identifying emotion words that are grounded in the theoretical framework of the works of Watson, Clark, and Tellegen (1988), Thompson (2007), and Diener et al. (2010). We use a basket of 69 English-extracted emotion words suggested to capture affect in the delivery of Positive and Negative Affect Schedules Extended (PANAS-X), International Positive and Negative Affect Schedule Short-Form (I-PANAS-SF), Scale of Positive and Negative Experience (SPANE), and various other studies including Engelen et al. (2006), Kahn et al. (2007), Dodd and Danforth (2010), Ford et al. (2018), Algan et al. (2019) and Boyd et al. (2022).

After selecting the abovementioned 69 words, we refined the list for the UK by testing the correlation of each word with True Happiness and retained only those that showed a statistically significant correlation. To further narrow the selection, we applied eXtreme Gradient Boosting (XGBoost), a machine learning algorithm, to rank the words based on their "gains," indicating the most important predictors of True Happiness. Next, we determine the weighting of the words (features) using the estimated coefficients of each word derived from predicting True Happiness using ElasticNet, a machine learning regression algorithm.

To test the accuracy of our derived equation to estimate UK happiness, we compare it to the UK's True Happiness measure. The results show a good fit (RMSE = 0.09), indicating that our index has a high level of accuracy in estimating happiness at the country level.

We perform various additional robustness tests using different frequencies of the extracted data (time-invariance) and unseen datasets (other periods). We also validate our Google Trends™ happiness index against another Big Data measure, namely the World Health Organisation's Early AI-supported Response with Social Listening (EARS). The dataset does not measure happiness but mental health and loneliness; therefore, a significant negative correlation will indicate our index's robustness.

To address our second research question, we use the same derived equation containing the same selected words and weights for the UK translated into Dutch to predict happiness. After applying the equation and estimating happiness in the Netherlands for 2011 and 2020, we evaluated the fit against the happiness measure from the Dutch Time Use survey data. We find different results considering different time periods. The results show a good fit for 2011 (RMSE = 0.08) and a weaker fit for 2020 data (RMSE = 0.43). However, the Time Use data quality weakened over time with fewer respondents and fewer observations, which might have contributed to the weaker fit.

To address our last research question, we re-estimate our happiness index, including country-specific Dutch words and validate it using Dutch Time Use data. The error metrics (RMSE = 0.05) for 2011 indicate a marginally better fit than using the primary selected emotion words and their weights and a significantly better fit for 2020. We find an overlap of the most important words in the UK and the Netherlands, but adding a few country-specific words improves the fit statistics.

Therefore, our results show that we achieve an acceptable fit using the same basket of words across countries, demonstrating our model's adaptability and scalability to different cultural and linguistic contexts. However, we can improve the fit by including country-specific emotion words to accurately estimate happiness levels from information-seeking query data extracted continuously from Google Trends™.

By achieving our aim of developing and validating a real-time happiness index, we offer governments and other stakeholders access to timely and relevant information about the mood of their citizens, which is applicable for decisive decision-making at significantly lower costs than survey data with a possibility to automate, to some extent, the process of measuring happiness.

The rest of the paper is structured as follows. The next section contains our theoretical framework for the emotion words and provides a literature review pertaining to studies that measured real-time happiness or subjective well-being. The data, selected variables and methodology are discussed in section 3. The results follow in section 4, while the paper concludes in section 5.

2 Literature Review

This section first discusses the theoretical framework for our emotion words and studies that explored the use of affect words. The next section discusses studies that developed real-time measures for happiness or subjective well-being using social media or search engines.

2.1 Measuring Affect

2.1.1 Theoretical Framework

Watson, Clark and Tellegen (1988) developed the Positive and Negative Affect Schedules (PANAS). The original PANAS included a 20-item bidimensional scale, which were broadly independent and discrete dimensions of affect rather than polar opposites on a continuum. These words were "Interested", "Distressed", "Excited", "Upset", "Strong", "Guilty", "Scared", "Hostile", "Enthusiastic", "Proud", "Irritable", "Alert", "Ashamed", "Inspired", "Nervous", "Determined", "Attentive", "Jittery", "Active" and "Afraid".

Watson and Clark (1994) expanded the original PANAS, known as PANAS-X, by creating a 60-item measure, which now expanded the two original higher-order scales to include 11 specific affects: "Fear", "Sadness", "Guilt", "Hostility", "Shyness", "Fatigue", "Surprise", "Joviality", "Self-Assurance", "Attentiveness", and "Serenity". The PANAS-X, therefore, measures mood at two different levels.

The original set of 20 words in the PANAS has faced some criticism. Validation studies using structural equation modelling, such as those by Crawford and Henry (2004), indicate that the most accurate models emerge when correlations are allowed between errors of items within the same word clusters from which the PANAS was initially developed (refer to Zevon & Tellegen, 1982, for word-cluster descriptors). These item covariances suggest a high degree of redundancy among certain PANAS items with similar meanings. Crawford and Henry's (2004) analysis demonstrated that the 10 items of the Negative Affect (NA) scale form five pairs with significant covariance: "distressed" and "upset," "guilty" and "ashamed," "scared" and "afraid," "nervous" and "jittery," and "hostile" and "irritable." Similarly, the Positive Affect (PA) scale's ten items cluster into four groups with shared variance. Two groups contain three items each: "interested," "alert," and "attentive," and "excited," "enthusiastic," and "inspired." The remaining two groups are formed by two pairs: "proud" and "determined," and "strong" and "active." These findings suggest that reducing the number of PANAS items may be possible without significantly compromising the PA and NA scales' content coverage or internal consistency.

Kercher (1992) created a shortened version of the original PANAS, reducing it to 10 items: "Excited," "Enthusiastic," "Alert," "Inspired," "Determined," "Distressed," "Upset," "Scared," "Nervous," and "Afraid." However, Mackinnon et al. (1999) noted that Kercher's abbreviated version included items with high covariance, which undermines content validity while artificially increasing internal consistency reliability. Furthermore, the full PANAS and Kercher's (1992) shortened version contained items with unclear or ambiguous meanings to both native and non-native English speakers from outside North America. For instance, many non-native English speakers do not understand the term "jittery," which is considered colloquial in many dictionaries. Additionally, Mackinnon et al. (1999) found that even among native English speakers, the item "excited" in Kercher's short form correlated significantly with both Positive Affect (PA) and Negative Affect (NA), suggesting it carries dual meanings for some.

To address redundancy issues and ambiguous meanings in different research contexts, Thompson (2007) developed the I-PANAS-SF (International Positive and Negative Affect Schedule Short Form). This new version was validated across various national, cultural, and occupational groups, demonstrating strong psychometric properties, including cross-sample stability, internal consistency, temporal stability, cross-cultural factorial invariance, and convergent and criterion-related validity. The I-PANAS-SF uses the question stem "Thinking about yourself and how you normally feel, to what extent do you generally feel:" and includes 5 Positive Affect and 5 Negative Affect items: "Upset," "Hostile," "Alert," "Ashamed," "Inspired," "Nervous," "Determined," "Attentive," "Afraid," and "Active."

2.1.2 Studies Focusing on Affect Words

A study applicable to the aim of this paper is Jovanović et al. (2022), who used the SPANE (Scale of Positive and Negative Experience), created by Diener et al. (2010), to determine its cross-cultural utility by measuring the invariance of the SPANE. The SPANE consists of 12 items designed to assess how often positive (SPANE-P subscale) and negative (SPANE-N subscale) emotions are experienced. It was created to address the limitations and challenges identified in previous emotion measurement tools, such as the PANAS. These include "Positive", "Negative", "Good", "Bad", "Pleasant", "Unpleasant", "Happy", "Sad", "Afraid", "Joyful", "Angry" and "Contented". Jovanović et al. (2022) focused on 13 countries: the United States, Turkey, Spain, Serbia, Portugal, Poland, Japan, Italy, India, Greece, Germany, Colombia and China. They found that SPANE's positive emotion terms, "positive", "good", "pleasant", "happy", "joyful", "contented", and general negative emotion terms, "negative" and "unpleasant", could be suitable for studies on emotions and well-being in a cross-cultural project.

Apart from the above, we also relied on the LIWC-22 dictionary (Boyd et al., 2022), a text analysis tool designed to assess language's psychological, social, and linguistic dimensions. LIWC-22 builds on previous versions by expanding its dictionary, refining its algorithms, and enhancing its usability for psychology, linguistics, and other social sciences researchers. LIWC was also validated by Kahn et al. (2007) as a valid tool for measuring emotional expression.

Other studies we used to identify emotion words include Engelen et al. (2006), which validated the Dutch versions of the PANAS, confirming their reliability and applicability for Dutch-speaking populations. The authors emphasised the importance of cultural adaptation when translating psychological measures. While the PANAS was originally developed in English, the study found that culturally sensitive translations retain the measure's effectiveness and ensure it remains meaningful across different linguistic and cultural contexts. We also considered the study done by Banerjee (2018), who used Google search data to study the patterns in public interest and concern related to the "internet", "anxiety", and "happiness", exploring how they are interrelated and vary across different countries and cultures. The author found that search volume data indicate significant interest in understanding how these topics connect to daily life, personal well-being, and mental health and that searches for "internet" often correlate with searches for "anxiety" and "happiness." This suggests a potential link between internet use and psychological states, where people might be using the internet both as a tool for coping with anxiety and as a means to seek or understand happiness.

Our last three studies, Dodds and Danforth (2010), Algan et al. (2019) and Ford et al. (2018), are discussed in sections 2.2 and 2.3 since they relied on measuring happiness or subjective well-being using Twitter or Google Trends^{TM}.

2.2 Measuring Happiness or Subjective Well-being Using Twitter

The pioneering research conducted by Dodd and Danforth (2010), Iacus et al. (2015, 2022), and Greyling and Rossouw (2019) are essential for measuring subjective well-being or happiness using Big Data sources like Twitter and will be further discussed below.

The Hedonometer was one of the pioneering tools developed to measure happiness in almost real-time using Big Data. Initiated by Dodds and Danforth (2010) at the end of 2008, the project tracked daily happiness levels, creating a continuous time series from late 2008 to May 2023 when Twitter (now X) suspended all academic licenses (refer to Dodds et al. (2011) for the foundational study). To begin, the authors merged the 5,000 most common words from four sources: Twitter posts, articles from the New York Times, Google Books, and Music lyrics. After merging the words, they are left with a composite set of around 10,000 unique words. They then used Amazon's Mechanical Turk to rate each word's happiness on a scale from 0 (unhappy) to 10 (happy), with "laughter" scoring the highest at 8.5 and "terrorist" scoring the lowest at 1.3. To construct the Hedonometer, they bin all the tweets extracted daily; however, only words recognised as English were included. The bin includes, on average, 200 million words extracted worldwide daily. Using a bag-of-words methodology, they assign a happiness score to each word, which is then averaged to produce a daily happiness index.

Iacus et al. (2015, 2022) were among the first to create a composite index of subjective and perceived well-being, encompassing various aspects of both individual and collective life. However, the measure was developed based on a priori-defined dataset without real-time predictive power. They developed their Subjective Well-being Index (SWBI) by applying an Integrated Sentiment Analysis (iSA) to tweets from Italy (starting in 2012) and Japan (beginning in 2015). The SWBI consists of eight components that reflect three distinct areas of well-being: social well-being, personal well-being, and well-being at work, with the final score being the average of these components. For instance 2015, Italy's SWBI averaged 48.7, while Japan's averaged 54.4. Carpi et al. (2022) used Random Forest and ElasticNet to analyse the impact of external factors such as data on the spread of COVID-19, economic indicators, air quality, internet searches and mobility data on the SWBI for Japan and Italy. Among others, they found that data on the spread of COVID-19, such as the number of deaths and cases, were more important for the SWBI in Japan than in Italy. Air quality was only relevant to the SWBI in Italy, whereas economic indicators were more relevant to the SWBI in Japan.

The *Gross National Happiness.today* project was launched by Greyling and Rossouw (2019) to determine national happiness levels (evaluative mood) in near real-time during different social, economic, and political events. They created their high-frequency daily time-series data by extracting live tweets and applying natural language processing (NLP) to analyse the sentiment. The sentiment analysis uses a lexicon-based approach incorporating tools like TextBlob, VADER, Sentiment140, and NRC, classifying tweets as positive, negative, or neutral. A balancing formula calculates a happiness score, which is averaged hourly and daily to provide near real-time time-series data. The scores range from 0 to 10, with 5 representing a neutral state, neither happy nor unhappy. In 2020, the project expanded to measure eight distinct emotions based on Plutchik's (1980) wheel of emotions, generating daily time-series data for each emotion. This project was also temporarily halted after Twitter (now X) suspended all academic licenses.

2.3 Measuring Mental Health, Life Satisfaction and Subjective Well-being Using Google Trends™

Murtin and Salomon-Ermel (2024) used Google Trends™ data to nowcast national average subjective well-being estimates for 38 OECD countries since 2010. To train their nowcasting models, they collected a large sample of time series from Google Trends™, covering 158 topics and 914 categories of searches chosen based on their relevance according to the American Time Use Survey, the OECD Well-being framework, and the domains of life satisfaction in happiness studies. The authors created a condensed version of the Google Trends™ dataset, where multiple time series were aggregated into subtopics based on the facets of the OECD Better Life Index. They derived 42 composite variables representing different dimensions of well-being. Their control variables included GDP per capita (with constant prices and purchasing power parity), the inflation rate, and the participation rate for individuals between 15 and 64 years old. They utilised large, customised micro-databases to improve model training on thoroughly pre-processed Google Trends™ data. Their findings related to life satisfaction indicated that the most accurate one-year-ahead predictions were achieved using a meta-learning approach that integrates forecasts from an ElasticNet model (both with and without interactions), a Gradient-Boosted Tree, and a Multi-layer Perceptron. Consequently, for 38 countries from 2010 to 2020, the out-of-sample prediction of average subjective well-being achieved an R^2 of 0.830.

Algan et al.'s (2019) study investigated how changes in internet search volumes can model and estimate subjective well-being in the United States. The authors used data from Google Trends™ to analyse the relationship between search behaviours and well-being measures from Gallup Analytics, covering the period from 2008 to 2013. The study developed national and state-level models using search data condensed into composite categories (e.g., job search, civic engagement, healthy habits) that reflect different life dimensions. Both models showed high out-of-sample predictive accuracy and effectively captured well-being trends. Using stepwise regression, they found that searches related to job search, civic engagement, and healthy habits consistently predict well-being (Gallup's indicators for "life evaluation today", "life evaluation in 5 years", "happiness", "laugh", "learn" and "respect") across multiple datasets and models. Job search terms are generally associated with lower well-being, while searches about civic engagement and healthy habits correlate with higher well-being.

Brodeur et al. (2020) utilised Google Trends™ data to examine the impact of government-imposed lockdowns on mental health and well-being. Their findings revealed a negative effect, indicated by increased searches related to sadness, worry, and loneliness. Foa et al. (2022) used two years (2020-2021) of Google Trends™ data from six English-speaking countries, along with weekly data from YouGov's Great Britain Mood Tracker Poll, to explore changes in subjective well-being throughout the COVID-19 pandemic. Using Google search terms such as "stress", "boredom", "frustration", "sadness", "loneliness", "feeling scared" ("fear"), "apathy", "happiness", "contentment", "energy", "inspiration" ("artistic inspiration"), and "optimism", they found that across the population, a decrease in affect tend to be associated with pandemic outbreaks. Furthermore, they found that while negative affect increased at the onset of lockdown, countries typically revert to baseline levels within three weeks at most, after which a net decrease in negative affect is observed.

In their study, Ford et al. (2018) used 15 terms from the PANAS-X to test the extent to which aggregated scores of emotion-related Google search queries are valid as indicators of subjective well-being at the US state and metro area levels. The selected terms included "afraid", "anxiety", "depression", "fatigue", "fear", "lonely", "nervous", "scared", "sleepy", "stress", "tired", "energetic", "enthusiastic", "happy", and "strong". The authors examined correlations between Google search scores and Gallup-Healthways measures of experienced negative emotions, namely "stress", "worry", "anger", and "sadness", as well as a composite measure combining these four emotions. They found that "afraid" was the most robust search term as it had significant associations with all the Gallup-Heatlthways indicators. Searches for "fear" were positively related to Gallup's "stress", "worry", and general negative affect. Searches for "scared", "lonely", and "nervous" were related to Gallup indicators of "anger" and "sadness", although "nervous" was also related to general negative affect. Interestingly, search scores for "depression" and "stress" were negatively related to Gallup "anger", while search scores for "anxiety" did not correlate with any Gallup items. The searches related to low arousal, e.g., "tired" and "fatigue", showed no relationships with Gallup indicators.

3 Data and Methodology

3.1 Data

3.1.1 Primary Dataset – Big Data Using Google TrendsTM

Google Trends TM is an open data service provided by Google Inc., allowing researchers to explore the temporal patterns of internet search activity based on specific keywords. It offers access to a single metric: the Relative Search Volume (RSV), a standardised measure reflecting search activity relative to the chosen time frame and geographic region. The RSV values range from 0 to 100, enabling comparisons of search volume trends across different queries, time periods, and locations (Houghton et al., 2023). Data from Google TrendsTM excludes certain data from searches. First, it excludes topics of interest where the interest is very low. Google TrendsTM only analyses data for popular terms, so search terms with low volume appear as 0 for a given time period. Second, Google TrendsTM excludes duplicate searches. It removes repeated searches by the same user within a short time frame to enhance overall accuracy. Lastly, Google TrendsTM excludes special characters by filtering out queries with apostrophes and other special characters.

Working with Google Trends TM data has certain limitations. First, Google TrendsTM data loses predictive power over time due to changes in search activity and the interface of Google Search itself (for example, auto-suggestion). For our purposes, this means that we need to regularly review the affect words in Table 4 and make any necessary adjustments to the weighting within our regression models. This will also allow us to periodically update and revise our index, which is also important for the model to incorporate social changes that could necessitate a re-weighting of the components.

Second, the search volume value on any given day cannot be directly compared across different terms because each term is normalised to its own maximum value. To resolve this issue, we standardise all search volumes to have a mean of zero and a standard deviation of one, focusing on changes in volume within each search term rather than relative differences between terms.

Third, Google Trends™ data presents estimation challenges because it does not provide raw search volumes; instead, it represents the proportion of total searches containing a specific keyword over a given period, normalised so that the highest value is 100. This normalisation affects the data interpretation in two main ways: first, the values directly obtained from Google Trends™ can be complex to interpret since they are influenced by both the search volume of the keyword and the overall search activity. Second, values on a given day cannot be compared between different terms, as each is scaled to its maximum. We standardised all search volumes to focus on within-term changes to address these issues rather than comparing absolute search volumes.

Fourth, while Google Trends™ provides a valuable real-time measure of public interest, its data is inherently influenced by demographic biases in search behaviour. Internet access, digital literacy, and platform preferences vary significantly across age groups, socioeconomic statuses, and geographic regions, leading to the potential overrepresentation of certain populations. This is a limitation in general when using digital data sources; therefore, any data derived from these sources must be validated against survey data, which is not susceptible to this limitation. In the current study, we validate our measures against survey measures of happiness.

Additionally, relying on Google Trends™ data to measure real-time happiness presents a risk over which researchers have no control, primarily due to the unpredictability of large tech platforms in maintaining their services. Previously, major players like Google, Meta and Twitter (now X) have abruptly discontinued services (e.g., Google Mobility Maps), restricted API access (e.g., Instagram, which impacted numerous applications, including popular dating apps like Tinder and Hinge, as well as the journaling app Day One), modified algorithms without warning (as frequently seen with Meta's advertising platform) or shut down its API service altogether (Twitter). Such changes can disrupt research or analytics that depend on consistent data streams. Therefore, we continuously explore alternative digital resources to derive well-being measures.

Apart from the above, it is important to consider that there may be a disconnect between survey-based measures and online search behaviours, which should be considered when selecting and using survey-derived keywords in the context of internet search data. This implies that if we consider positive and negative emotions, we must also consider the search actions that will be taken if these emotions are experienced, thereby necessitating a wider scope of words than only emotion words.

These words could also likely include those that are searched for when people experience these emotions and can vary from words related to entertainment (e.g., "movies," "Netflix," "music"), social relationships (e.g., "family," "friends"), well-being (e.g., "thankfulness"), to actions taken (e.g., suicide) when people experience negative emotions.

For example, in the Gallup data (Helliwell et al., 2024), the measurement of positive affect (similar to our happiness measure) is defined by the average of three positive affect measures: "laughter", "enjoyment", and "doing interesting things". These measures are obtained from responses to the following three questions: "Did you smile or laugh a lot yesterday?", "Did you experience the following feelings during A LOT OF THE DAY yesterday?" How about Enjoyment?" and "Did you learn or do something interesting yesterday?". While these questions capture aspects of positive affect, the associated keywords may not necessarily reflect how people search for related content online.

Information seeking is a fundamental human drive, arising when people face perplexing situations or become aware of gaps in their knowledge, kindling a desire to fill those voids of understanding (Ford et al., 2018). However, because querying search engines is a goal-oriented solitary activity, there's a fundamental difference in how bridging informational gaps must be examined relative to conventional surveys. To effectively explore positive emotions through search engine queries, survey questions probing affirmative sentiments must be translated into information-seeking terms that capture the nuances of each distinct emotion when submitted to a platform. Indeed, second-person-centric survey questions (e.g., "Did you do or learn something interesting yesterday?") will be translated into instances of their first-person-centric equivalences ("hobbies to explore nearby?"; "movies screening today?"; "concert nearby today?").

For example, using Gallup data, you will not have a Google query such as "Did you smile or laugh yesterday?". Rather, the query could be, "Which movies are showing?” which relates to activities undertaken when experiencing positive emotions. This must be considered when establishing emotion keywords to extract from Google Trends™ to measure experienced happiness.

Information-seeking queries are more likely to include negative emotion words when a person is experiencing an emotion reflecting negative affect, and the search is seeking information on how the negative affect can be changed to a positive emotion (Ford et al., 2018).

In the Gallup data, negative affect is defined and measured as the average of three negative affect measures. They are "worry", "sadness", and "anger", respectively, the responses to "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Worry?" "Did you experience the following feelings during A LOT

OF THE DAY yesterday? How about Sadness?" and "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Anger?" (Helliwell et al., 2024).

Therefore, we need to consider "queries seeking information on experienced negative affects" to measure happiness with information-seeking queries to construct a happiness index. If we translate the negative affect (emotions) to an information-seeking question, it could be "What can I do to decrease my anger/worries/sadness? What can I do to minimise the experience of negative emotions? or What can I do to be happy?", thereby maximising positive emotions.

To compile our Google Trends™ dataset, we data-mined the emotion-specific Google queries according to the list of 69 words (see Table 4) derived from the literature and theory (sections 2.1 to 2.3) for the UK for the period from January 2011 to December 2023 at a daily frequency. We extracted the data using the Gtrends library in R.

Table 4: The 69 words extracted to establish those with the highest correlation with True Happiness.

great	joke	attentive	cry	punish	wellbeing	angry
party	joy	inspired	dead	reject	well-being	cancer
game	love	active	depressed	sad	suicide	divorce
comedy	music	alone	disease	sick	sleep	hopeless
friendship	pleasure	abuse	fear	stress	sadness	pain
fun	win	afraid	hate	tired	boredom	weak
good	movie	anxiety	headache	worry	depression	joyful
happy	song	anxious	kill	wrong	loneliness	contented
health	friend	bad	lonely	panic	ashamed	determined
hope	alert	crime	nervous	upset	unpleasant	

In the initial construction phase, we collapse the daily data to weekly data, summing the observations for the period coinciding with the availability of ONS data per week, from 5 January 2020 to 1 October 2023, giving us 196 observations.

3.1.2 Secondary Datasets – Survey and Big Data

3.1.2.1 Survey Data

We considered samples of high-frequency survey data that measure happiness dynamics as a source to validate our new Google Trends™ happiness index. High-frequency survey data measuring happiness is scarce and mainly limited to the US[1] and the UK, although we also have access to the Dutch Time Use survey data. Therefore, the availability of the UK and Dutch data directed our choice of countries in our exploratory analyses.

To address our first research question, we chose the UK's Office for National Statistics (ONS) survey data for the period from 2020 to the end of 2023, with a weekly frequency, which is publicly available. Additionally, as the main language in the UK is English, it allows us to concentrate on only one language.

Specifically, the data we used from ONS forms part of the UK tracking their progress across 10 domains of national well-being, including personal well-being, relationships, and health. Within these 10 domains are 44 indicators of national well-being, including people rating their life satisfaction, happiness, anxiety and whether their lives are worthwhile. We rely on the question where adults aged 16 years and over were asked to rate how happy they felt yesterday on a scale from 0 to 10, where 0 was "not at all" and 10 was "completely" (ONS, 2023). From this point forward, it is referred to as True Happiness.

For the ONS data, we have 196 observations measuring happiness from 5 January 2020 to 1 October 2023, which we use as the outcome variable in training our models to predict happiness. We include the entire period in our analysis to maximise the number of observations. While we acknowledge the presence of a structural break in the data during the COVID-19 pandemic, we opted to use the whole time series to enhance the likelihood of achieving an accurate model fit. This decision allows us to test our methodology and derive an equation to estimate happiness with high accuracy.

To address our second research question, we use happiness aggregated at the daily level from the Dutch Time Use survey (Bakker et al., 2020) as an outcome variable to explore whether the same basket of emotion words selected for the UK translated into a different language (Dutch) with the same derived weights can also successfully estimate happiness in another country. To test the robustness of our index, we calculated the error metrics between the Dutch Time Use data, which spans the period 2011 to 2021 and our derived happiness measure for the Netherlands. However, there are limitations to the Dutch Time Use survey data, which can lead to a decrease in fit statistics in later years. Since 2020, there have been high levels of missingness per day or very few observations, limiting the representativeness of the data.

[1] We did not choose the US since the Gallup American Time Use Survey data is not available unless we incur significant costs.

3.1.2.2 Big Data

We also use Big Data to validate our UK happiness index in the form of the World Health Organisation's (WHO) Early AI-supported Response with Social Listening (EARS) daily dataset. Initially, EARS was used to show real-time information about how people were talking about COVID-19 online. The data was compiled so that the WHO could better manage the situation as the infodemic and pandemic evolved. Although it did not measure happiness, it did measure mental health and loneliness, which we determined from section 2 should have some relationship with happiness. Specifically, we use document 12, representing the number of documents per day for the category mental health and document 33, representing the number of documents per day for the category loneliness.

3.2 Methodology

In this section, we explain the methodology followed to derive our Google Trends™ happiness equation to estimate happiness at a country level.

3.2.1 Correlation to Decrease Number of Words

After we mined all 69 words from Google Trends™ (section 3.1.1), we tested correlations between weekly measures of the UK's True Happiness and the positive and negative affect emotion words extracted from Google Trends™. We remind the reader that the basket of words includes positive and negative words since positive and negative affect are not the inverse of one another but rather independent and discrete dimensions of affect. At the same time, a person can experience both emotions, which complicates the measure of happiness using information-seeking queries. Suppose the negative affect words are highly correlated with True Happiness (negatively). In that case, we can assume that it has an inverse relationship to True Happiness and is closely related to the measure of positive affect (although the relationship is negative). It can ultimately be used in constructing a happiness index since negative affect drives action, i.e., people are more likely to search for solutions, causes, or validation when experiencing distress (e.g., "how to deal with anxiety"). In addition to capturing positive emotions, we also include search actions which will be undertaken when people experience positive emotions – for example, entertainment-related searches (e.g., movies, Netflix, music), social relationship-related words (e.g., family, friends, friendships) and well-being-related words (e.g., thankfulness).

We selected words statistically significantly correlated to True Happiness, leaving us with 42 words. Since we aim to predict happiness using the most important emotion keywords, we continue the process of decreasing the number of words, training a model using eXtreme Gradient Boosting.

3.2.2 eXtreme Gradient Boosting (XGBoost)

To identify the most important words (features) for predicting happiness, we employed the eXtreme Gradient Boosting (XGBoost) algorithm, a highly efficient and scalable machine learning method that implements gradient boosting for decision trees. XGBoost operates within the gradient boosting framework, where models are developed sequentially, and each new model is trained to correct the errors of its predecessor. This iterative process continues until a robust predictive model is achieved. XGBoost is specifically designed for high performance and speed, utilising optimisation techniques that enable parallel and distributed computing, making it well-suited for large datasets. It also incorporates regularisation methods (L1 and L2) to mitigate the risk of overfitting.

XGBoost has proven to be more accurate than other methods. For example, Abdurrahim et al. (2020) compared various predictive modelling algorithms and found that XGBoost achieved the highest accuracy score when compared to methods like random forest, decision trees, naive Bayes classifier, and logistic regression.

Therefore, our XGBoost model is defined in equation (1) as:

$$F_M(x) = F_0 + v\beta_1 T_1(x) + v\beta_2 T_2(x) + \cdots + v\beta_M T_M(x) \qquad (1)$$

Where M is the number of iterations. The gradient boosting model is a weighted $(B_1 \dots \beta_M)$ linear combination of simple models $(T_1 \dots T_M)$. $F_M(x)$ is the True Happiness, also known as the target variable, measured weekly from 2020 to 2023, with 196 observations. From September 2020, data was only reported every second week. We imputed the data for this period. The independent variables (features) are the 42 words selected through the correlation exercise.

Model evaluation uses metrics to analyse the model's performance, i.e., how well the model generalises future predictions. Machine learning metrics include Accuracy, Precision, Recall and F1 score in classification problems with a discrete, often binary outcome variable. However, we make use of the Mean Absolute Error (MAE), the Mean Squared Error (MSE), and the Root Mean Square Error (RMSE) since our outcome variable is continuous.

3.2.3 Weighting (ElasticNet)

Basic econometric methods are not an option to predict True Happiness as we have many independent variables (features) that are highly correlated (multicollinearity). Working with weekly data of the ONS only available for the period 2020 to 2023 limits our observations to 196. Therefore, using an estimation technique such as OLS to determine which words significantly predict True Happiness was not an option, as the low number of observations means we have insufficient degrees of freedom, and we are challenged by multicollinearity.

Therefore, we turned to ElasticNet, which is a regularised regression ML technique that incorporates both L1 (Lasso) and L2 (Ridge) regularisation penalties into its objective function. It is designed to handle many features in smaller datasets. By combining the L1 and L2 penalties, ElasticNet can achieve both feature selection and parameter shrinkage, making it particularly useful in scenarios with highly correlated predictors (highly correlated words, as in the present study). In addition, it mitigates the risk of overfitting by shrinking less important coefficients and potentially setting some to zero (like Lasso); however, it does not fully eliminate the risk, particularly when working with small datasets as in the current study.

The success of ElasticNet depends heavily on the choice of its hyperparameters, namely:

Alpha: The mixing parameter between Lasso (L1) and Ridge (L2) penalties. Alpha = 1 corresponds to Lasso, while Alpha = 0 corresponds to Ridge. In our analysis, we used the default Alpha parameter of 0.5.

Lambda (or L1_ratio): The regularisation strength. Higher values mean more regularisation.

Due to the small sample size, K-fold cross-validation is crucial to ensure the model generalises well to new data. This technique involves partitioning the data into K subsets and then iteratively training the model on K-1 subsets while using the remaining subset for testing.

In addition to using K-fold cross-validation to mitigate the potential of overfitting, we also introduced other measures, as seen in section 3.2.4. Nonetheless, we are aware of the potential of reduced generalisability of the models when applied to unseen data when interpreting results.

We use the estimated coefficients of the features from ElasticNet as weights to derive our happiness index, which estimates country-level happiness.

3.2.4 Steps Taken to Guard Against Overfitting

Due to the risk of overfitting when using relatively smaller datasets, we included the following steps in both the XGBoost and the ElasticNet models. First, in terms of the XGBoost, we used i) a strict feature selection, limiting the number of features to only those words statistically significant to the survey happiness measure, i.e., “True Happiness"; ii) randomised sample selection throughout; iii) small tree depth of 3 to limit the complexity, which normally improves generalisation; iv) specified an early stop criteria by monitoring the RMSE; if there were no improvements after five iterations, the training process was stopped. In terms of ElasticNet, as mentioned above, it is adapted explicitly to small sample sizes to prevent overfitting with its regularisation (Lambda and Alpha) and adds penalties to the coefficients (shrinking the coefficients). Furthermore, the features included in the ElasticNet models underwent a two-stage selection process whereby only statistically significant words were included in the XGBoost – and using the XGBoost, the features were further reduced to only include the most important features. The strictly selected features limit the risk of including noise in the estimation. We used k-fold cross-validation to assess model performance and ensure it generalises well.

For both the XGBoost and ElasticNet, we monitored the performance metrics (RMSE) of the training and test sets. The metrics were very similar, indicating a good fit – if there were large gaps, it might have indicated overfitting.

3.2.5 Robustness Checks

Lastly, we rely on unseen data to test the robustness of our derived Google Trends™ happiness index. To test the time invariance of our derived index, we applied our happiness equation to quarterly ONS data to determine if the trends captured using weekly data and quarterly are similar.

Additionally, we test the robustness of our Google Trends™ happiness index using Big Data. Here, we relied on daily data from EARS: document 12, representing the number of documents per day for the mental health category and document 33, representing the number of documents per day for the category loneliness, to test the correlations between our daily derived happiness index and these measures.

4 Results and Analysis

4.1 Constructing a Happiness Index Using Google Trends™ for the UK

4.1.1 XGBoost Initial Model on Search Terms' Relative Importance

To determine the most important words from the 42 words identified in section 3.2.1, we use XGBoost (section 3.2.2). The outcome variable (Label) is the UK's True Happiness (happiness as reported in the ONS survey data). We use the regression option within the XGBoost algorithm since True Happiness is a continuous variable.

We start by randomly splitting the data into a training and testing dataset with an 80:20 split on all data, with the evaluation done on the unseen testing data. We used the random split method to ensure that both datasets represent the overall distribution.

To train the model, we initially used all the default settings of the parameters of the XGBoost algorithm. Next, we predict True Happiness, evaluate the model according to the fit metrics, and refine the parameters to optimise the model's performance. We found the optimal tree depth was three[2]. We set the number of iterations to 100, with a termination clause added to stop the algorithm if the RMSE does not decrease after 5 iterations.

The RMSE evaluating the fit reached its lowest value of 0.11268 at the 32nd boosting cycle, after which the training stopped due to the "early stop" criteria of 5 if the RMSE did not improve.

The evaluation metrics for our XGBoost model show that all measures of fit reveal small errors, indicating a good-fitting model. For the XGBoost, the MSE is 0.013, the MAE is 0.083, and the RMSE is 0.113.

Following the results from the XGBoost model, those words with gains of more than 0.01 were retained, leaving us with 26 words.

From the 26 words, we found that "sad" was the most important word with a gain of 0.2571, followed by "headache" (gain of 0.1657), "depressed" (gain of 0.0547) and "music" (gain of 0.0423). It is interesting to note that negative emotion words indeed are significant predictors of True Happiness. This agrees with our earlier discussion that using information-seeking Google queries will most likely lead to finding the negative queries important. Positive words that are important predictors include "well-being" (gains = 0.0401), "love" (gains= 0.0366) and "great" (gains= 0.0236).

4.1.2 ElasticNet Weighting and Aggregation

As mentioned in section 3.2.3, we used an ElasticNet linear regression algorithm, a machine learning approach, to estimate the coefficients. Once the coefficients are determined, we use these as weights in the equation to estimate the happiness levels in countries.

To train the ElasticNet model, we randomly split the data into a training and testing dataset with an 80:20 split on all data. We initially started by using all the default parameters for ElasticNet with an Alpha and Lambda of 0.5. After conducting the 5-fold cross-validation process, we identified that the optimal (best) Lambda was 0.000193946, which minimised prediction error and indicated a moderate level of regularisation, which is in line with the complexity of the dataset.

The evaluation metrics indicate a good fit with an MAE of 0.0690, an MSE of 0.0117, and an RMSE of 0.0938, indicating that our model predicts True Happiness well. The scatter plot in Figure 3 confirms the good fit of the Predicted Happiness versus True Happiness scores.

[2] In XGBoost, tree depth refers to the maximum depth of individual trees in the ensemble. A depth of three means that each tree in the boosting process was limited to three levels of splits, optimising the trade-off between model complexity and generalisation while mitigating overfitting.

Figure 3. Predicted Happiness vs True Happiness for the UK

Therefore, the generic equation to estimate happiness in the UK is as follows:

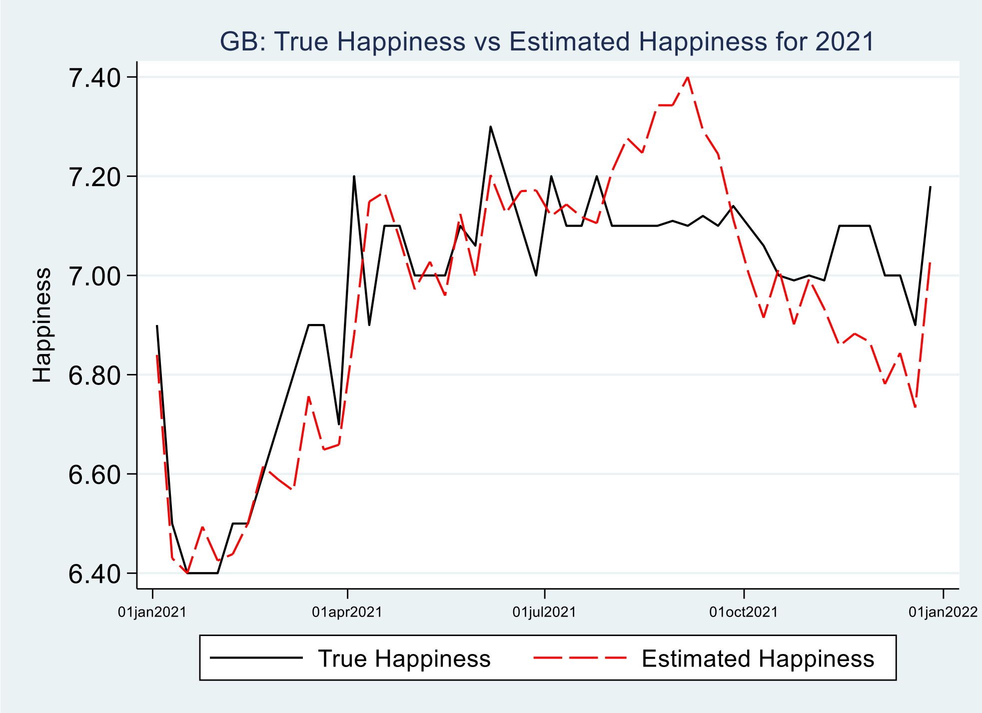
$$GNH\_GT = \beta_0 + \alpha_1 * \beta_1 + \cdots\cdots\cdots + \alpha_{26} * \beta_{26} \qquad (2)$$

Where $\alpha_{1_1} \cdots \alpha_{26}$ represents the 26 words as determined by our XGBoost model to be the most important words in predicting the outcome variable, True Happiness, and $\beta_1 \cdots \beta_{26}$ represent the weights as determined by the coefficient results from the ElasticNet linear regression model.

For example, in the newly derived equation to estimate happiness, the weights for "sad", "headache", and "depressed" are -0.1324, -0.1230 and -0.0228, respectively.

Applying the newly derived equation, we estimate happiness in the UK. Figure 4 shows the *Estimated Happiness* versus True Happiness (ONS weekly survey data). Evaluating the fit, the RMSE is 0.0940, which indicates a very good fit.

Figure 4. Estimated Happiness vs True Happiness from the UK ONS data



4.1.3 Results from Robustness Checks and Validation Exercise

As mentioned in section 3.2.4, we test the time invariance of our Google Trends™ happiness index by applying our derived equation (2) to quarterly ONS data. We observe that the trends captured using weekly data are also reflected in the quarterly data. Table 5 shows the correlation between the estimated happiness indices using weekly and quarterly data, which is strong and significant at 0.7000 (p=0.000).

Additionally, we validate our Google Trends™ happiness index using Big Data. Here, we tested the correlation between our derived happiness index and daily data from EARS: document 12, representing the number of documents per day for the mental health category and document 33, representing the number of documents per day for the category loneliness. Table 5 shows statistically significant and negative correlations of -0.31 (mental health) and -0.25 (loneliness). The negative correlations are as expected.

Table 5. Correlation of Google Trends™ happiness index using quarterly ONS data and high-frequency data from EARS.

Measure	Estimated Happiness
Estimated Happiness	1
ONS Quarterly Happiness	0.7000 ***
EARS – doc 12 (mental health)	-0.3121 ***
EARS – doc 33 (Loneliness)	-0.2425***

Source: Authors' own calculations.

Considering the correlation results in Table 5, we are confident that our equation yields consistent results regardless of the data frequency (e.g., weekly or quarterly), and it is robust when validated against other well-being measures.

4.2 Constructing a Happiness Index Using Google Trends™ for the Netherlands

4.2.1 Estimating Happiness in the Netherlands Using the UK-derived Equation

This section reports the results of our second research objective. Here, we explore whether the same basket of words used in the UK index translated into a different language (Dutch) with the same derived weights can also successfully estimate happiness in another country. As mentioned in section 3.1.2, we translated the basket of 26 words determined by our XGBoost model into Dutch and applied our equation to the extracted Dutch words. To test the validity of our *Estimated Happiness*, we correlate it with True Happiness as recorded in the Dutch Time Use survey data (Bakker et al., 2020).

Table 6. Correlation of Google Trends™ happiness index in Dutch correlated to Dutch Time Use survey data.

Measure	Estimated Happiness
Estimated Happiness	1
Dutch Time Use Survey Happiness - 2011	0.5742***
Dutch Time Use Survey Happiness - 2020	0.2589***

Source: Authors' own calculations.

Table 6 shows that the estimated happiness using the Dutch equivalent of our English words and weights (equation to estimate happiness for the UK) is statistically significantly correlated to happiness recorded in the Dutch Time Use Survey. In 2011, the correlation was strong, at 0.5742, though it performed much weaker in 2020, with the correlation being 0.2589.

Figures 5 and 6 show the *Estimated Happiness* against True Happiness measured by the Dutch Time Use survey data. The fit statistics show an RMSE of 0.08 for 2011; however, the fit weakens considerably with an RMSE of 0.43 for 2020. The results are similar to those revealed using correlation analysis. The result of a weaker fit in 2020 is surprising as we expected the estimated happiness to show a better fit for 2020, as the period coincides with that used to derive the UK equation. However, a likely explanation is Dutch Time Use data quality, as it has deteriorated over time, with significantly fewer respondents and many missing observations since 2020.

Figure 5. Estimated Happiness vs True Happiness from the Netherlands Dutch Time Use survey data (2011)

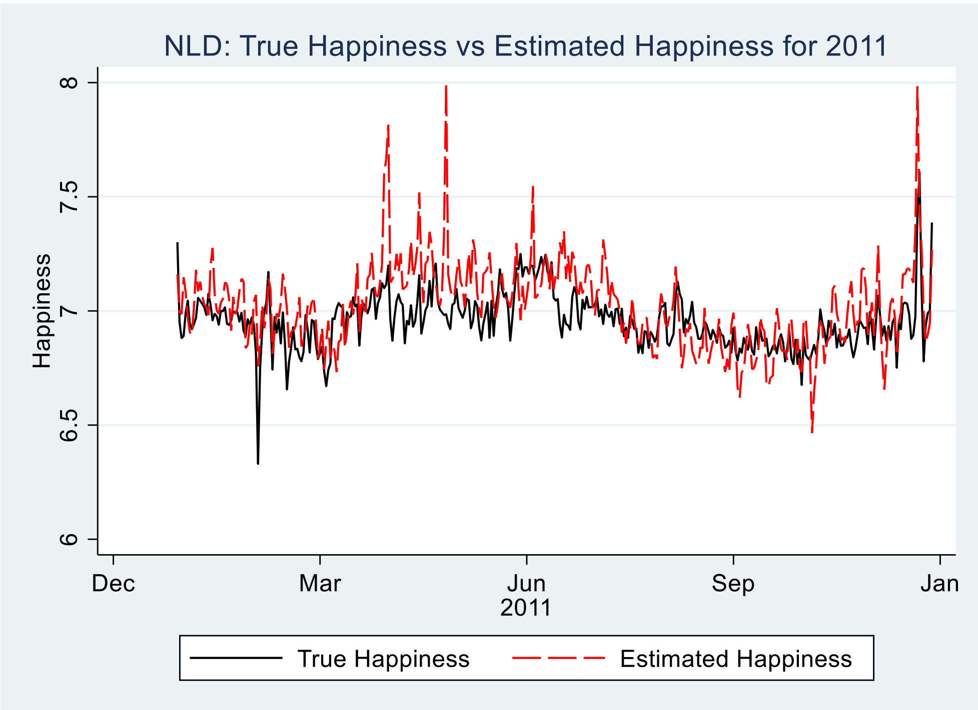
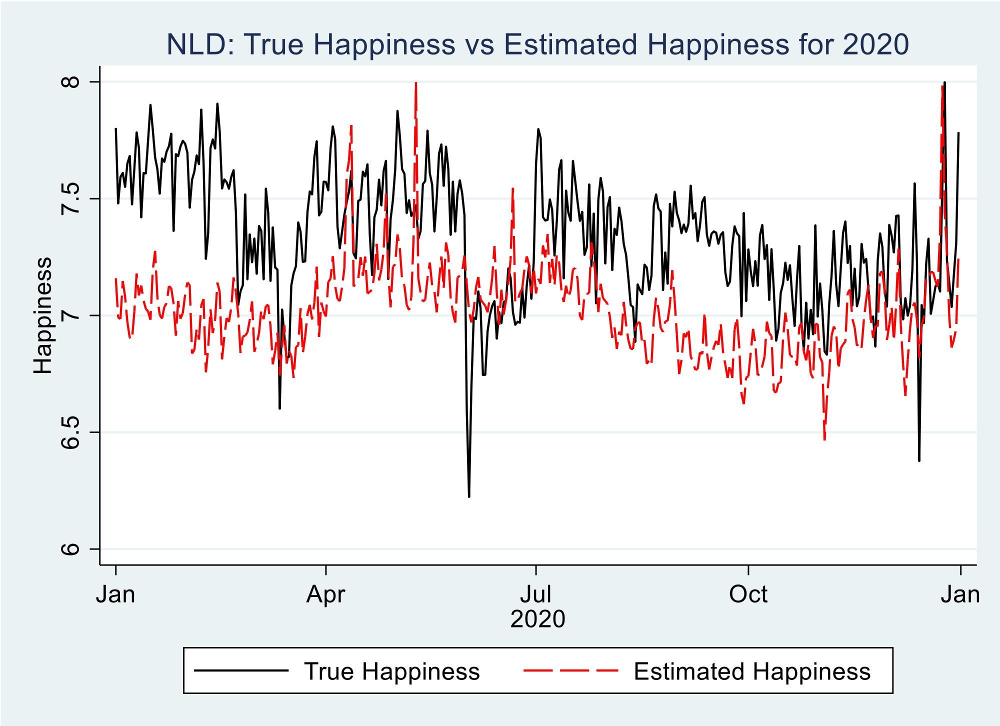


Figure 6. Estimated Happiness vs True Happiness from the Netherlands Dutch Time Use survey data (2020)



Considering our results, we re-estimate our happiness equation for the Netherlands in the next section by incorporating country-specific words. To predict True Happiness using our machine learning algorithms in the Netherlands, we use the Dutch Time Use Survey data for the period 2011 and 2012 (during these two years, the quality of the Time Use data was good).

4.2.2 Constructing a Happiness Index Using Google Trends™ for the Netherlands, Including Country-specific Emotion Words

This section reports the results of our third research objective. Here, we use the same set of words as those we used for the UK but add relevant country-specific words such as "perfect" and "fijne" to reflect emotion words used in Dutch. We started our analyses with the same initial 69 emotion words (refer to Table 4) translated into Dutch, adding the country-specific words. Subsequently, we used correlation analysis to establish which of the extracted words were significantly correlated to the Dutch Time Use survey's happiness measure (Bakker et al., 2020) and reduced the initial basket of words to 47 words.

To determine the most important words from the 47 words identified above, we use XGBoost (section 3.2.2). The outcome variable (Label) is the Netherlands' True Happiness (Dutch Time Use survey's happiness measure). Using XGBoost, we follow the method explained in section 4.1.1 by randomly splitting the data into a training and testing dataset with an 80:20 split on all data, with the evaluation done on the unseen testing data.

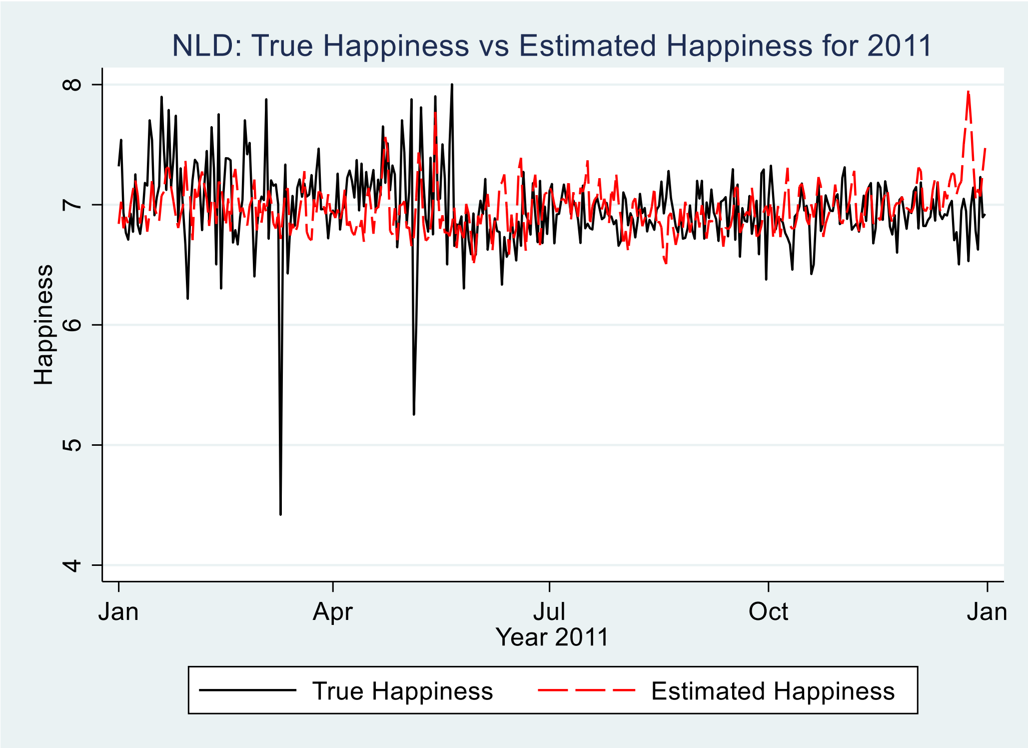
The evaluation metric for our XGBoost model reveals a small error, indicating a good-fitting model. For the XGBoost, the RMSE is 0.0501. Following the results from the XGBoost model, those words with gains of more than 0.01 were retained, leaving us with 23 words. The most important words included, among others, "fijne" (gain = 0.1837), "kanker" (gain = 0.1304), "hoofdpijn" (gain = 0.0785), and "dood" (gain =0.0638).

We use the most important words as features to train the ElasticNet regression model to predict True Happiness in the Netherlands. The evaluation metrics demonstrate strong predictive performance with an MAE of 0.0345, an MSE of 0.0025, and an RMSE of 0.0504, indicating that our model predicts True Happiness well.

We use the estimated coefficients to weight the features (words) to derive the happiness equation for the Netherlands. Figure 7 shows the True Happiness (from the Dutch time use survey data) and the Estimated Happiness for the year 2011.

A visual inspection suggests a good fit. To evaluate the accuracy of our Estimated Happiness versus True Happiness, we calculated the RMSE. The RMSE for True Happiness versus Estimated Happiness is 0.0504, indicating a smaller error compared to the initial derived happiness equation, which had an RMSE of 0.0823, where no country-specific words were included (see section 4.2.1). Furthermore, the fit is markedly better than the one we attained for 2020, with an RMSE=0.43.

Figure 7. Estimated Happiness vs True Happiness from the Netherlands Dutch Time Use Survey data with country-specific emotion words



Therefore, we can conclude that adding country-specific words decreases errors and improves the accuracy of happiness estimations using information-seeking query data extracted continuously from Google Trends™.

5 Conclusions

In this paper, we constructed and validated a real-time happiness index using search query data based on emotion keywords, which we extracted from Google Trends™, representing the first of its kind to our knowledge. Our Google Trends™ happiness index for the UK combines 26 words, each with its own weight as determined by our ElasticNet linear regression machine learning model.

We initially started with carefully curated words suggested to capture positive and negative affect. We extracted the 69 words using Google Trends™ from the UK and correlated those to the weekly UK True Happiness measure obtained from the ONS data. Words significantly correlated to the happiness score were selected for further analysis. We were left with 42 words, which subsequently decreased to 26, given our results from the XGBoost model that determined the most important words (features) in predicting True Happiness obtained from the UK's ONS data (outcome variable). Subsequently, we used the ElasticNet linear regression machine learning model to estimate the coefficients of each of the 26 words in predicting happiness. These coefficients were then applied as weights in our equation to estimate happiness.

To test the time invariance of our Google Trends™ happiness index for the UK, we applied our derived equation to quarterly ONS data. Additionally, we validated our Google Trends™ happiness index using Big Data in the form of EARS: document 12 represents the number of documents per day for the category mental health, and document 33 represents the number of documents per day for the category loneliness. Considering the correlation results, we are confident that our equation yields consistent results regardless of the data frequency (e.g., weekly or quarterly), and it is robust when validated against other well-being measures.

We used data from the Netherlands to explore whether the same basket of words translated into a different language (Dutch) with the same derived weights can also successfully estimate happiness. We translated our 26 English words into Dutch and applied our derived happiness equation. Then, we correlated it with the True Happiness measure from the Dutch Time Use survey and found a statistically significantly strong relationship in 2011 and a weaker relationship in 2020. We also plotted our *Estimated Happiness* versus True Happiness in 2011 and 2020 and calculated the respective RMSEs. The results showed a good fit for 2011 (RMSE = 0.08) and a weaker fit for 2020 (RMSE = 0.43). A plausible reason is the quality of the Time Use data, which weakened over time with fewer respondents and fewer observations.

Lastly, we followed the same methodology we used for the UK to derive a happiness equation. Here, we used the 69 identified emotion words and added country-specific words. We determined the most important words using XGBoost. We used those words in an ElasticNet algorithm to estimate the coefficient of each word and subsequently applied them as weights in our happiness equation for the Netherlands. Using the derived equation, we estimated Dutch happiness. To test the accuracy of our estimated happiness, we compared it to True Happiness. We calculated the RMSE, which is 0.05, indicating that although emotion words show an overlap between countries, including country-specified words can improve happiness estimations.

Therefore, we successfully showed that information-seeking queries extracted using Google Trends™ can be used to estimate happiness and construct a near real-time happiness index.

The result of our study provides several practical initiatives. First, governments and policymakers can leverage the real-time insights provided by the Google Trends™ happiness index to monitor national mood fluctuations and respond accordingly to mitigate potentially violent and destructive outcomes such as violent riots. Second, since happiness is linked to productivity and economic performance, monitoring happiness trends in real-time can inform labour policies. For example, significant dips in happiness might indicate rising workplace dissatisfaction, burnout, or economic hardship, prompting governments to adjust workplace well-being initiatives or financial support programmes. Third, real-time tracking of happiness can help policymakers anticipate public reactions to major policy decisions. By analysing how past legislative changes, such as the COVID-19 mandates, impacted national happiness, decision-makers can better predict potential resistance to new policies and implement communication strategies to mitigate backlash. Lastly, our study has policy implications for international organisations such as the OECD and the United Nations that aim to measure global well-being. Policymakers should consider local linguistic and cultural variations when designing happiness-tracking frameworks.

It is important to acknowledge two key considerations related to our happiness measure. First, our Google Trends™ happiness index is time-sensitive, requiring intermittent review and confirmation of the selected emotion words to ensure that our derived happiness equation is still accurate in estimating a country's happiness. Moreover, while the composition of happiness equations may vary slightly across countries—each incorporating a few different emotion words—we maintain that these differences do not compromise the validity of the measure. Despite cultural variations in the expression of happiness, the concept itself is broadly recognised across societies. As such, while some individual terms may differ, all happiness equations ultimately converge on the same fundamental outcome: a meaningful representation of well-being.

Second, while we demonstrated the effectiveness of using Google Trends™ to measure real-time happiness in the UK and the Netherlands (English and Dutch), the generalisability of these findings to other significantly different linguistic and cultural contexts remains an open question. Languages differ in how they encode and express emotions, making direct translation of happiness-related terms difficult for capturing cultural nuances. Additionally, internet search behaviours vary across regions due to disparities in digital access, privacy concerns, and the use of alternative search engines. To enhance cross-cultural validity, future research should incorporate expert linguistic reviews to refine translated emotion terms. Computational techniques, such as multilingual word embeddings, could further improve scalability by identifying semantically equivalent emotion words across languages. A crowdsourced approach involving native speakers could also help refine emotion lexicons, ensuring broader applicability and consistency across different cultural contexts.

Our future research will endeavour to use our Google Trends™ methodology to construct and estimate indices of subjective well-being at a country level in real-time, including life satisfaction. Apart from extending our research agenda, we are committed to the continuous refining and adaptation of our methodology to ensure its robustness across diverse linguistic, cultural, and technological landscapes. We will continue to explore strategies to enhance the flexibility of our approach, including integrating dynamic keyword validation and employing human-in-the-loop validation processes to refine emotion lexicons. Additionally, we will expand our dataset to include a broader range of countries, which will allow us to understand better the influence of digital behaviour on our happiness measure. By actively iterating on our methods, we aim to strengthen the scalability, reliability, and universality of Google Trends™ as a tool for real-time well-being assessment.

References

Abdurrahim, Y., Ali, A. D., Sena, K., & Huseyin, U. (2020). Comparison of deep learning and traditional machine learning techniques for classification of pap smear images. *arXiv*, 2009.06366v1. Available from https://arxiv.org/pdf/2009.06366.pdf

Algan, Y., Murtin, F., Beasley, E., Higa, K., & Senik, C. (2019). Well-being through the lens of the internet. *PLOS ONE* 14(1), e0209562.

Bakker, A., Burger, M., van Haren, P., Oerlemans, W., & Veenhoven, R. (2020). Raise of happiness following raised awareness of how happy one feels: a follow-up of repeated users of the happiness indicator website. *International Journal of Applied Positive Psychology*, *5*, 153-187.

Banerjee, S. (2018). How Does the World Google the Internet, Anxiety, and Happiness? *Cyberpsychology, Behavior, and Social Networking, 21*(9), 569 – 574.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin. https://www.liwc.app

Brodeur, A., Clark, A. E., Fleche, S. & Powdthavee, N. (2020). Assessing the impact of the coronavirus lockdown on unhappiness, loneliness, and boredom using Google Trends. arXiv: 2004.12129. Available at https://ui.adsabs.harvard.edu/abs/2020arXiv200412129B/abstract

Bryson, A., Clark, A. E., Freeman, R. B., & Green, C. (2016). Share capitalism and worker well-being. *Labour Economics, 42*, 151–158.

Callegaro, M. & Yang, Y. (2018). The Role of Surveys in the Era of "Big Data". In Vannette D., Krosnick J. (eds) *The Palgrave Handbook of Survey Research*. Palgrave Macmillan, Cham.

Carammia, M., Iacus, S. M., & Wilkin, T. (2022). Forecasting asylum-related migration flows with machine learning and data at scale. *Scientific Reports, 12*(1), 1-16.

Carpi, T., Hino, A., Iacus, S. M., & Porro, G. (2022). The Impact of COVID-19 on Subjective Well-Being: Evidence from Twitter Data. *Journal of Data Science*, *21*(4), 761-780.

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 3*(Pt 3),245-65.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*, 143-156.

Dodds, P. S., & Danforth, C. M. (2010). Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. *Journal of Happiness Studies, 11,* 441–456.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A. & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLOS ONE, 6*(12), e26752.

Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., & Van Den Bergh, O. (2006). Verdere validering van de Positive and Negative Affect Schedule (PANAS) en vergelijking van twee Nederlandstalige versies [Further validation of the Positive and Negative Affect Schedule (PANAS) and comparison of two Dutch versions]. *Gedrag & Gezondheid: Tijdschrift voor Psychologie en Gezondheid, 34*(2), 89–102.

Foa, R. S., Fabian, M., & Gilbert, S. (2022). Subjective well-being during the 2020–21 global coronavirus pandemic: Evidence from high frequency time series data. *PLOS ONE, 17*(2), e0263570.

Ford, M. T., Jebb, A. T., Tay, L., & Diener, E. (2018). Internet Searches for Affect-Related Terms: An Indicator of Subjective Well-Being and Predictor of Health Outcomes across US States and Metro Areas. *Applied Psychology: Health and Well-being, 10*(1), 3–29.

Greyling, T., Rossouw, S., & AFSTEREO. (2019). *Gross National Happiness.today Index*. Available from http://gnh.today

Greyling, T. & Rossouw, S. (2022). Positive attitudes towards COVID-19 vaccines: A cross-country analysis. *PLOS ONE, 17*(3), 0264994.

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2024). *World Happiness Report 2024*. Appendix 1: Statistical Appendix for Chapter 2. University of Oxford: Wellbeing Research Centre.

Houghton, S., Boy, F., Bradley, A., James, R., Wardle, H., & Dymond, S. (2023). Tracking online searches for gambling activities and operators in the United Kingdom during the COVID-19 pandemic: A Google Trends™ analysis. *Journal of Behavioral Addictions*, *12*(4), 983-991.

Iacus, S. M., Porro, G., Salini, S., & Siletti, E. (2015). Social networks, happiness and health: From sentiment analysis to a multidimensional indicator of subjective well-being. *ArXiv*. Available from https://doi.org/10.48550/arXiv.1512.01569

Iacus, S. M., Porro, G., Salini, S., & Siletti, E. (2022). An Italian Composite Subjective Well-Being Index: The Voice of Twitter Users from 2012 to 2017. *Social Indicators Research, 161*, 471–489.

Jovanović, V., Joshanloo, M., Martín-Carbonell, M., Caudek, C., Espejo, B., Checa, I., Krasko, J., Kyriazos, T., Piotrowski, J., Rice, S. P. M., Junça Silva, A., Singh, K., Sumi, K., Tong, K. K., Yıldırım, M., & Żemojtel-Piotrowska, M. (2022). Measurement Invariance of the Scale of Positive and Negative Experience Across 13 Countries. *Assessment*, *29*(7), 1507–1521.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring Emotional Expression with the Linguistic Inquiry and Word Count Source. *The American Journal of Psychology, 120*(2), 263-286.

Kasser, T., & Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Personality and Social Psychology Bulletin, 22,* 280–287.

Kercher, K. (1992). Assessing subjective well-being in the old-old. The PANAS as a measure of orthogonal dimensions of positive and negative affect. *Research on Aging, 14*(2), 131–168.

Kim, E.S., Kubzansky, L.D., & Smith, J. (2015). Life satisfaction and use of preventive health care services. *Health Psychology*, *34*(7), 779– 782.

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the positive and negative affect schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences, 27*, 405–416.

Murtin, F., & Salomon-Ermel, M. (2024). Nowcasting subjective well-being with Google Trends: A meta-learning approach. OECD papers on well-being and inequalities. Working Paper No.27. Available from https://doi.org/10.1787/4ca48f7c-en

ONS (2023). Measures of National Well-being: Quality of life in the UK – May 2023. Available from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/measuringnationalwellbeingdomainsandmeasures

Piekalkiewicz, M. (2017). Why do economists study happiness? *Labour, 28*(3), 361–377.

Rossouw, S., & Greyling, T. (2024). "41: Big data and happiness". In Encyclopedia of Happiness, Quality of Life and Subjective Wellbeing. Cheltenham, UK: Edward

Elgar Publishing. Available from https://doi.org/10.4337/9781800889675.00051

Rossouw, S., & Greyling, T. (2020). Big Data and Happiness. Invited chapter for the Handbook of Labor, Human Resources and Population Economics. Edited by Klaus F. Zimmermann. Available from https://doi.org/10.1007/978-3-319-57365-6_183-1#DOI

Thompson, E. R. (2007). Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227–242.

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

Watson, D., & Clark, L. A. (1994). The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form. Available from https://iro.uiowa.edu/esploro/outputs/other/The-PANAS-X-Manual-for-the-Positive/9983557488402771

Williams, S., & Shiaw, W. T. (1999). Mood and organisational citizenship behavior: The effects of positive affect on employee organisational citizenship behavior intentions. *Journal of Psychology, 133,* 656–668.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic nomothetic analysis. *Journal of Personality and Social Psychology, 43*(1), 111–112.

Appendix A - The comparability of differently worded subjective well-being measures

Abstract

What and how we measure affects what we do. It might therefore be concerning that the wording of single-item subjective well-being (SWB) measures differs between commonly used surveys in the SWB literature. The aim of this study is to provide a better understanding of whether and how wording differences between SWB measures affects people's responses. Using experimental data from over 7,700 respondents across four experiments, our findings show that the wording of single-item happiness and life satisfaction measures has limited effects. While lower SWB and higher dispersion is occasionally observed in SWB questions with rarely used neutral or negative question tones, more common wording differences do not affect the predictors, means, and dispersion of SWB. The findings imply that while the lack of a uniform SWB measure is not ideal, it does not pose a significant threat to the credibility of findings from the SWB literature.

1 Introduction

The use of SWB metrics and the science on SWB is quickly expanding in the "beyond GDP" era. The most used SWB metrics are single-item self-report measures on life satisfaction and happiness. While most SWB measures have the same core elements and intended meaning, every major survey uses a differently worded SWB measure (see Table 7). The SWB measures differ in terms of question tone, scope indicators, and the wording of the scale (see Table 8). While a vast literature has examined and proven the credibility of SWB metrics (OECD, 2013), subjective measures, and thus also self-reports of SWB, are affected by context effects (Schwarz and Strack 1999). Even if question wording could cause context effects, the impact of wording differences in SWB measures has been largely neglected. Yet, while not examining question wording, a nascent study of Bjørnskov (2010) shows that slightly different SWB concepts can cause substantial differences in the levels and predictors of happiness.

While some wording differences between major surveys are minor (e.g., synonymous words), other differences may cause more substantial differences in how the respondent interprets and answers the question, such as the inclusion of a time scope indicator referring to the present (e.g., "nowadays"), the use of a bipolar versus unipolar scale, or what anchor is used in the question (i.e., the question tone). Most SWB measures have a positive question tone, meaning that the question itself refers to how "happy" or "satisfied" the respondent is. The anchoring, priming and framing literatures suggest that such words can serve as anchors and primes for respondents, meaning that they use it as a starting point in their thought process (Kahneman and Tversky 1974; Furnham and Boo 2011).

To the author's knowledge, no existing surveys are available measuring the same SWB concept with differently phrased single-item questions. We conducted four experiments. The goal of the first two experiments is to examine whether the means, dispersion, and predictors of respectively life satisfaction and happiness are sensitive to question tone. The third and fourth experiment replicate the same question tone differences in different cultural and linguistic contexts, examines additional differences in the wording of SWB measures beyond question tone, and explore potential mechanisms causing differences in results.

Table 7. SWB Measures in the Most Used Surveys in the SWB Literature

Survey	Question	Scale
Panel A: Life satisfaction		
OECD guidelines	Overall, how satisfied are you with life as a whole these days?	0 (not at all satisfied) -10 (completely satisfied)
WVS	All things considered, how satisfied are you with your life as a whole these days?	1 (completely dissatisfied) - 10 (completely satisfied)
ESS	All things considered, how satisfied are you with your life as a whole nowadays?	0 (extremely dissatisfied) - 10 (extremely satisfied)
SOEP	How satisfied are you with your life, all things considered?	0 (completely dissatisfied) - 10 (completely satisfied)
HILDA	All things considered, how satisfied are you with your life?	0 (totally dissatisfied) - 10 (totally satisfied)
BHP	How dissatisfied or satisfied are you with your life overall?	1 (not satisfied at all) - 7 (completely satisfied)
Latinobarometro	Generally speaking, would you say you are satisfied with your life?	1 (very satisfied) - 4 (not at all satisfied)
BRFSS	In general, how satisfied are you with your life?	1 (very satisfied) - 4 (very dissatisfied)
Eurobarometer	On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead?	1 (very satisfied) – 4 (not at all satisfied)

Panel B: Happiness		
OECD guidelines	Taking all things together, how happy would you say you are?	0 (not at all happy) - 10 (completely happy)
ESS	Taking all things together, how happy would you say you are?	0 (extremely unhappy) - 10 (extremely happy)
WVS	Taking all things together, would you say you are:	1 (very happy) - 4 (not at all happy)
GSS	Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy, or not too happy?	1 (very happy) - 3 (not too happy)
CGSS	Generally speaking, do you think your life is happy?	1 (very unhappy) - 5 (very happy)

Note: WVS=World Values Survey; ESS=European Social Survey; SOEP= German Socio-Economic Panel; HILDA= Household, Income and Labour Dynamics in Australia; BHP=British Household Panel; BRFSS=Behavioral Risk Factor Surveillance System (USA); GSS=General Social Survey (USA); CGSS=Chinese General Social Survey. “OECD guidelines” refers to the recommended question by the OECD (2013).

Table 8. Question wording differences (based on Table 7)

Question aspect	Wording
Panel A: Life satisfaction	
Question tone	positive (7); neutral (2)
Scope wording	all things considered (3); overall (2); generally speaking (1); in general (1); on the whole (1)
Place of scope wording	beginning (7); end (2)
Secondary scope wording	as a whole (3); none (6)
Time scope	these days (2); nowadays (1); none (6)
Scale label	numerical (6); verbal (3)
Scale type	unipolar (4); bipolar (5)
Scale extremes	completely (4); very (3); extremely (1); totally (1)
Panel B: Happiness	
Question tone	positive (3); neutral (2, of which 1 open)
Scope wording	taking all things together (3); Taken all together (1); generally speaking (1)
Place of scope wording	beginning (5)
Time scope	these days (1); none (4)
Scale label	numerical (2); verbal (3)
Scale type	unipolar (3); Bipolar (2)
Scale extremes	very (3); extremely (1); completely (1)

Note: This table provides an overview of question wording differences in the surveys included in Table 1. A positive question tone refers to questions asking how “satisfied” or “happy” the respondent is. A neutral score refers to questions with “dissatisfied or satisfied” (BHP), all answering options (Eurobarometer and GSS) and an open question without anchors (WVS).

2 Methodology

We conducted four experiments that were designed to have complementary strengths (and weaknesses) and that serve as robustness checks for each other. The characteristics and experimental conditions of the four experiments are presented in Tables 3 and 4. The four experiments feature 18 experimental conditions, 8 for life satisfaction and 10 for happiness. The experimental conditions are designed so that (1) the impact of a wording difference can be isolated by having maximum one wording difference compared to one other condition, and (2) the questions are where possible identical to existing survey questions. All experiments used a between-subjects design, meaning that respondents were randomly assigned to one experimental condition. Hence, respondents received and answered only one experimental question, except for experiment 3 where respondents answered two experimental questions, one for life satisfaction and one for happiness. Experiments 1 and 4 focused on life satisfaction, experiment 2 focused on happiness, and experiment 3 focused on both life satisfaction and happiness. While experiments 1 and 2 focused only on question tone, experiments 3 and 4 included more experimental conditions that also covered wording differences in terms of scope wording, time scope, scale type, and scale extremes as well as questions about mechanisms that may explain possible differences between conditions. We can examine the external validity of our findings by having experiments in four different languages, across different cultures, and with different populations (samples representative of a country's adult population vs. student samples).

Randomization was successful in all four experiments (see Appendix Tables A1-A5). Significant group differences at the 95% confidence level occurred for 0 out of 9 variables in experiment 1, 0 out of 12 variables in experiment 2, 1 (for life satisfaction) or 2 (for happiness) out of 28 variables in experiment 3, and 1 out of 28 variables in experiment 4. Nevertheless, to rule out that subtle group differences bias the results, the reported results in the main analysis are conditional on all covariates used in the randomization check. Unconditional results are reported in the appendix. To provide a clear overview of the results for each theme (e.g., question tone, question wording) across experiments, the results section is organized by theme rather than by individual experiment.

Table 9. Survey Characteristics

Experiment	1	2	3	4
N	421	1,611	1,043	4,669
Sample	Greek students (convenience sample)	Representative of the Dutch adult population	Dutch & international students (convenience sample)	Representative of the Japanese adult population
Survey language	Greek	Dutch	English	Japanese
Data source	Experimental question included in survey of Arampatzi et al. (2020)[1]	Experimental question included in survey of CentERpanel[2]	Own survey through university courses[3]	Experimental question included in survey on the relationship with family, co-workers, and community members
Experimental conditions	4	5	16 (2x8)	8
SWB focus	Life satisfaction	Happiness	Life satisfaction and happiness	Life satisfaction
Question wording focus	Question tone	Question tone	Question tone Scope wording Scale wording	Question tone Scope wording Scale wording
Questions on mechanisms	No	No	Yes	Yes

Position of experimental question in survey	End of survey	End of survey	Beginning of survey	Middle of survey
Survey mode	Online, confidential	Online, confidential	Online, anonymous	Online, confidential

[1] Lecturers at various Greek universities invited students to voluntarily participate in the survey.
[2] The CentERpanel is a household panel that is maintained by CentERdata, a research institute affiliated with Tilburg University. The survey was completed by 78% of the panel members. The experimental question featured in a survey about lottery play (Burger et al. 2020). While SWB questions are not a regular part of this panel, the respondents are experienced in completing surveys.
[3] Students completed the questionnaire in preparation for an introductory lecture about SWB at the beginning of a course on the economics of well-being or a Massive Open Online Course (MOOC) on critical thinking. While English was not the mother tongue of most students, they had advanced language proficiency in English given that the courses were offered in English. Unlike the other experiments, the survey of experiment 3 was specifically designed for this study instead of being part of a multi-purpose survey.

Table 10. Experimental Conditions

Condition	Question	Response scale	Experiment 1	2	3	4
Life satisfaction						
Satisfied[1]	All things considered, how satisfied are you with your life as a whole these days?	0 (completely dissatisfied) - 10 (completely satisfied)	✓		✓	✓
Dissatisfied or satisfied[2]	All things considered, how dissatisfied or satisfied are you with your life as a whole these days?	“”	✓		✓	✓
Open[3]	All things considered, how do you feel about your life as a whole these days?	“”	✓		✓	✓

Dissatisfied[4]	All things considered, how dissatisfied are you with your life as a whole these days?	“”	✓		✓	✓
Satisfied (short)[5]	All things considered, how satisfied are you with your life?	“”			✓	✓
Satisfied (overall)[6]	Overall, how satisfied are you with life as a whole these days?	“”			✓	✓
Satisfied (overall; unipolar)[7]	Overall, how satisfied are you with life as a whole these days?	0 (not at all satisfied) - 10 (completely satisfied)			✓	✓
Satisfied (short; totally)[8]	All things considered, how satisfied are you with your life?	0 (totally dissatisfied) - 10 (totally satisfied)			✓	✓
Happiness						
Happy[9]	Taking all things together, how happy would you say you are?	0 (extremely unhappy) - 10 (extremely happy)[19]		✓	✓	
Happy or unhappy[10]	Taking all things together, how happy or unhappy would you say you are?	“”		✓	✓	
Unhappy or happy[11]	Taking all things together, how unhappy or happy would you say you are?	“”		✓		
Open[12]	Taking all things together, would you say you are?	“”		✓		

Unhappy[13]	Taking all things together, how unhappy would you say you are?	“”	✓	✓
Happy (unipolar)[14]	Taking all things together, how happy would you say you are?	0 (not at all happy) - 10 (completely happy)		✓
Happy (short; unipolar)[15]	Overall, how happy do you feel?	“”		✓
Open (verbal; short)[16]	Taking all things together, would you say you are:	1 (very happy) - 4 (not at all happy)		✓
Open (verbal)[17]	Taken all together, how would you say things are these days? Would you say that you are …?	“”		✓
Open (verbal; bipolar)[18]	Taken all together, how would you say things are these days? Would you say that you are …?	1 (very happy) - 4 (very unhappy)		✓

[1] Measure from the World Values Survey. This is used here as the base measure for life satisfaction.
[2-4] Deviate from the base measure in question tone. Conditions 2 and 3 have a neutral tone, condition 4 has a negative tone.
[5] Question (not scale) from HILDA. Deviates from the base measure in scope wording.
[6] Recommended question (not scale) by the OECD guidelines. Deviates from the base measure in scope wording.
[7] Recommended measure by the OECD guidelines. Deviates from condition 6 in scale wording.
[8] Measure from HILDA. Deviates from condition 5 in scale wording.
[9] Measure from the European Social Survey. This is used here as the base measure for happiness.
[10-13] Deviate from the base measure in question tone. Conditions 10-12 have a neutral tone, condition 13 has a negative tone.
[14] Recommended measure by the OECD guidelines. Deviates from the base measure in scale wording.
[15] Deviates from condition 14 in scope wording and question length.
[16] Measure from the World Values Survey. Deviates from condition 17 in scope wording and question length.
[17] Question (not scale) from the General Social Survey. Deviates from condition 16 in scope wording and question length.
[18] Measure from the General Social Survey. Deviates from condition 17 in scale wording.

[19] Experiment 2 used a 1-7 scale to alleviate possible consistency bias from having answered non-experimental questions on life satisfaction and mood using 1-10 scales at the beginning of the survey.

3 Results

3.1 Question Tone

In the results section, a 95% confidence interval is used to report statistically significant differences. Table 5 shows for all four experiments the conditional means and dispersion of life satisfaction and happiness for the experimental conditions differing by question tone.

3.1.1 Means

For life satisfaction, a one-way ANCOVA reveals that the null hypothesis of equal group means is rejected in the Japanese experiment (p=0.00), but not in the Greek and student experiments (p=0.11 in both cases). Post-hoc tests identified significant mean differences in life satisfaction in 1 out of 6 pairwise comparisons in the Greek and student experiments. In these experiments, mean life satisfaction is higher in the positively framed question *satisfied* condition compared to the neutrally framed *open* question. In the Japanese experiment, mean life satisfaction is also highest in the positively framed *satisfied* condition, but not lowest in the open condition. Specifically, 3 out of 6 pairwise comparisons show significant differences in the Japanese experiment, with life satisfaction being higher in the positively framed *satisfied* condition compared to the negatively framed *dissatisfied* condition and the neutrally framed *dissatisfied or satisfied* condition. Moreover, life satisfaction is higher in the open condition compared to the negatively framed *dissatisfied* condition. Across experiments, the differences ranged approximately between 0.2 and 0.4 points.

For happiness, the null hypothesis is rejected in both experiments (p=0.03 in the Dutch experiment and p=0.00 in the student experiment). Post-hoc tests identified significant mean happiness differences in 3 out of 10 pairwise comparisons in the Dutch experiment. Consistent with the results for life satisfaction, mean happiness is lower in the open-ended question compared to the positively framed 'happy' condition and the neutrally framed *happy or unhappy* and *unhappy of happy* conditions. The magnitude of the differences is modest, approximately 0.1 points on the 7-point scale. In the student experiment, which contains no open-ended question, 2 out of 3 pairwise comparisons show differences. Unlike the Dutch experiment, mean happiness is lower in the negatively framed "unhappy" condition compared to the other conditions. The magnitude of the differences is large, approximately 0.7 to 0.8 points on the 11-point scale.

Overall, combining all life satisfaction and happiness experiments, mean SWB is higher in positively framed questions compared to open questions in 3 out of 4 cases, higher compared to negatively framed questions in 2 out of 5 cases, and higher compared to non-open neutrally framed questions in 1 out of 6 cases. Similarly, mean SWB is higher in the non-open neutral questions compared to the open neutral questions in 3 out of 4 cases, and higher compared to the negatively framed questions in 1 of out 5 cases.

2.1.2 Dispersion

A Levene's test shows significant differences in dispersion in each experiment, except for the Dutch experiment. Across experiments, conditions with lower mean SWB tend to show higher dispersion. This is expected since the mean scores are above the mid-point of the 0-10 or 1-7 scales and are left-skewed due to a ceiling effect, with the right tail cut off at the scale maximum. Put differently, dispersion is higher at lower means because the right tail is less constrained by the scale maximum. Generally, dispersion does not differ between conditions with similar means. There are two exceptions to these patterns: In the Dutch experiment, the open condition does not have higher dispersion despite a lower mean; conversely, in the student experiment, the *dissatisfied* condition has higher dispersion despite a similar mean. Overall, the positively framed questions show lower dispersion than the open questions in 2 out of 4 cases, lower dispersion than the negatively framed questions in 3 out of 5 cases, and lower dispersion than non-open neutrally framed questions in 1 out of 6 cases. Similarly, the non-open neutrally framed questions show lower dispersion than the neutral open questions in 2 out of 4 cases (but higher dispersion in 1 case) and lower dispersion than the negatively framed question in 2 out of 5 cases. Similar patterns for means and dispersion are observed in the unconditional results (see Table A6).

Table 11. Question Tone- Means and Dispersion

	Experiment 1: Greece		**Experiment 2: Netherlands**		**Experiment 3: Students**		**Experiment 4: Japan**	
Condition	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Panel A: Life satisfaction								
Satisfied	7.17	0.89			7.27	0.85	5.86	1.12
Dissatisfied or satisfied	7.17	0.96			7.07	0.94	5.67	1.32
Open	6.87	1.10			6.90	1.21	5.80	1.18
Dissatisfied	7.14	0.84			7.13	1.37	5.54	1.46
Panel B: Happiness								
Happy			5.34	0.55	7.14	0.89		
Happy or unhappy			5.33	0.51	7.05	0.94		

Unhappy or happy			5.31	0.55				
Open			5.22	0.55				
Unhappy			5.29	0.57	6.32	1.78		

Note: Means and standard deviations (SD) are adjusted for differences in all covariates used in the randomization check of each experiment (supplemented by age^2 in experiments 2 and 4 to capture the common U-shaped relationship between SWB and age).

2.1.3 Predictors

To examine whether question tone affect the determinants of SWB, the bivariate relationship between SWB and a battery of socio-demographic, attitudinal, and personality covariates is estimated for each question tone condition. The full results for each experiment are reported in Tables A7-A8. Pairwise comparisons are used to examine whether bivariate relationships differ across conditions. The expected percentage of pairwise comparisons showing significant differences by chance (type I errors) is 5% at the 95% confidence threshold. The percentage of bivariate relationships showing differences across conditions ranges between 0% and 6% across conditions. Since the percentages are below or close to 5%, the observed differences across conditions are likely to be observed by chance. Zooming in on which relationships differ suggests that the differences are indeed non-systematic. First, the relationships do not differ more often when comparing conditions with diverging means and dispersion. For example, the open and positively framed conditions tend to differ the most in means and dispersion, but the bivariate relationships do not differ more than when making other pairwise comparisons. Second, there are no consistent patterns in which covariates differ between conditions across experiments. For example, social trust differs between the open and negatively framed conditions in experiment 1, but not in experiment 3. Finally, based on framing theory, one may expect that less positively (or more negatively) framed SWB questions relate more strongly to undesirable situations (e.g. worry and financial difficulty). Yet, no systematic differences are observed in how covariates capturing undesirable situations relate to the differently framed SWB questions. A robustness check where all variables are simultaneously included shows similar results.

2.1.4 Mechanisms

Experiment 3 allows for testing five possible mechanisms that could cause differences in results for the life satisfaction question, in particular the observed lower mean life satisfaction in the open condition and the higher dispersion in the open and negatively framed condition in experiment 3. The examined mechanisms are: (1) the main life domains considered by respondents when answering the question, (2) the perceived happiness norm, (3) how respondents interpreted the numerical scale, (4) the difficulty of answering the question, and (5) the response duration (see Table A9 for the exact measures). Table 12 shows the differences between conditions in the tested mechanisms. No differences are observed in 3 of the 5 explored mechanisms: the scale interpretation, perceived norm, and domain score of focus areas. This suggests that the tone of the question does not affect how respondents use the scale, their perception of people's average life satisfaction, and their level of satisfaction with the domains on which they focus. Regarding the other 2 mechanisms, respondents (a) consider the *satisfied* condition more difficult to answer than the *dissatisfied or satisfied* condition, (b) focus more on their environment in the *dissatisfied* condition than in the *dissatisfied or satisfied* condition, and (c) focus more on their family in the *open* condition than in the *satisfied* and *dissatisfied or satisfied* conditions. Yet, these differences cannot explain the lower mean life satisfaction in the open condition and the higher dispersion in the open and negatively framed condition. Overall, the interpretation of the measure and the thoughts provoked by the measures have limited sensitivity to question tone.

Table 12. Mechanisms

	(1) Satisfied	(2) Dissatisfied or satisfied	(3) Open	(4) Dissatisfied	(5) Significant differences
Focus: financial (%)[1]	26	22	26	28	
Focus: health (%)	43	38	41	37	
Focus: achievements (%)	46	37	41	41	
Focus: family (%)	48	54	66	55	3>1+2
Focus: work/study (%)	53	52	54	53	
Focus: safety (%)	17	18	16	17	
Focus: environment (%)	31	20	24	33	4>2
Domain score of focus area[2]	6.8	7.3	7.2	6.9	
Perceived norm	6.3	6.3	6.2	6.3	
Scale interpretation: A bit satisfied	5.7	5.8	5.7	5.8	
Scale interpretation: A bit dissatisfied	4.1	4.1	3.9	4.2	

Response duration[3]	11.4	11.8	10.8	12.6	
Question difficulty	1.7	1.5	1.6	1.6	1>2

[1] The domain percentages reflect the percentage of respondents considering the domain “very important” for choosing their answer.
[2] This score is calculated by a person’s average domain satisfaction on domains they considered “very important” for choosing their answer. Hence, a lower score means that domains a respondent was less satisfied with was considered more when answering the life satisfaction question.
[3] Medians are reported for response duration and differences calculated using a Dunn’s test to minimize effects of outliers.

2.2 Other Question and Scale Wording Differences

Table 13 shows the conditional means and dispersion of life satisfaction and happiness for the experimental conditions with other question and scale wording differences (see Table A10 for similar results using unconditional means). Results regarding the predictors and potential mechanisms are reported in Tables A11-A13.

2.2.1 Scope and Wording and Question Length

No significant differences in means or dispersion are observed in the following cases: (1) removing the secondary scope and time scope indicator “as a whole these days” (*short* vs. *satisfied* condition), (2) rephrasing the primary scope indicator as “overall” instead of “all things considered” (*overall* vs. *satisfied* condition), (3) rephrasing the primary scope indicator to “overall” instead of “taking all things together” and shortening the question by replacing “would you say you are” with “do you feel” (*unipolar* vs. *short; unipolar* condition), and (4) removing the time scope indicator “these days” and the secondary sentence “Would you say that you are …?” (*verbal* vs. *verbal; short* condition). No systematic differences are observed regarding the predictors (see Tables A11-12). Regarding the mechanism tests for life satisfaction, the *short* question is considered less difficult to answer than the *satisfied* condition, while respondents focused more on domains they are more satisfied with in the *overall* condition compared to the *satisfied* condition (see Table A13). However, these differences have not translated in differences in the means, dispersion and predictors of life satisfaction. The results suggest that SWB results are not sensitive to these types of scope wording and the shortening of questions.

2.2.2 Answer Scales

No significant differences in means or dispersion are observed when changing the quantity indicator of the answer scale from ‘completely’ to ‘totally’ (*short* vs. *short; totally* conditions). Two comparisons are made to assess sensitivity to unipolar versus bipolar scales using 11-items scales: the *overall* vs. *overall (unipolar)* condition for life satisfaction and the *unipolar* vs. *short; unipolar* condition for happiness. Both comparisons reveal no significant differences in means and dispersion. Finally, the unipolar verbal scale from the World Values Survey is compared to the bipolar verbal scale from the General Social Survey (*open (verbal)* vs. the *open (verbal; bipolar)* condition). These conditions also show no significant differences in means and dispersion. It should be noted, however, that mean life satisfaction did differ between the *Open (verbal; short)* and *Open (verbal; bipolar)* conditions, which differ both in question and scale wording. Regarding all these comparisons, the predictors did not systematically differ across conditions and the mechanism tests did not show differences.

Overall, only 1 out of 16 comparisons that can be made in Table 13 shows a significant mean difference (*open (verbal)* vs. *the open (verbal; bipolar)* condition). But this comparison is not an isolated difference (differs both in question and scale wording) and the difference is likely caused by the different verbal scale labels. The dispersion did not differ between any of the conditions. Moreover, we do not observe more differences in correlates than would be found by chance (between 4 and 7,5%; see Tables) and the mechanism tests did not show substantial differences. These findings show that SWB results have limited sensitivity to differences in question wording and scale wording other than question tone.

Table 13. Other Question and Scale Wording Differences- Means and Dispersion

	Experiment 3		**Experiment 4**	
Condition	*Mean*	*SD*	*Mean*	*SD*
Panel A: Life satisfaction				
Satisfied	7.27	0.85	5.86	1.12
Satisfied (short)	7.12	1.06	5.83	1.09
Satisfied (overall)	7.13	1.03	5.86	1.14
Satisfied (overall; unipolar)	7.06	0.82	5.84	1.18
Satisfied (short; totally)	7.21	1.05	5.81	1.23
Panel B: Happiness				
Happy	7.14	0.89		
Happy (unipolar)	7.18	0.90		
Happy (short; unipolar)	7.35	0.96		
Open (verbal; short)	3.17	0.45		

Open (verbal)	3.13	0.51	
Open (verbal; bipolar)	3.03	0.45	

Note: Adjusted means and standard deviations (SD) reported as in Table 11.

3 Preliminary Conclusions

The sensitivity of SWB outcomes to question wording appears to be limited.

Shortening questions, including removing secondary scope indicators like "as a whole" and "these days," or replacing phrases such as "all things considered" with "overall" or "would you say you are" with "do you feel," did not significantly impact the results.

Scale labels also exhibited little sensitivity, showing no notable differences between bipolar and unipolar scales or various quantity indicators for scale extremes in numerical scales. However, verbal scale labels did demonstrate a potential effect on the reported means.

Regarding question tone, the commonly used positively framed questions have in some cases higher means and lower dispersion compared to neutrally or negatively framed SWB questions, particularly when compared to rarely used open question frames. Despite these differences, SWB correlates were not systematically influenced by question tone.

Our findings imply that variations in question wording have a minor impact on the SWB literature. While we welcome efforts to minimize question wording differences, question wording differences do not present a significant threat to the credibility of SWB literature. Notably, given that most leading surveys in the SWB literature use positive or non-open neutral frames, the comparability of findings in the SWB literature is not substantially affected by question tone differences. However, the lower SWB scores when using an open or negative framing suggests that the current predominantly positively framed measures capture upper bounds of people’s true SWB. Furthermore, our findings suggest that using shorter SWB questions does not compromise results.

References

Arampatzi, E., Burger, M., Stavropoulos, S., & Tay, L. (2020). The role of positive expectations for resilience to adverse events: Subjective well-being before, during and after the Greek bailout referendum. Journal of Happiness Studies, 21, 965-995.

Bjørnskov, C. (2010). How comparable are the Gallup World Poll life satisfaction data?. Journal of happiness Studies, 11, 41-60.

Cantril, H. (1965). The pattern of human concerns. New Brunswick, NJ: Rutgers University Press.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. Journal of Personality Assessment, 49(1), 71-75.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. Journal of Socio-Economics, 40(1), 35-42.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37(6), 504-528.

OECD (2013), OECD Guidelines on Measuring Subjective Well-being, OECD Publishing, Paris,

Richins, M. L. (2004). The material values scale: Measurement properties and development of a short form. Journal of Consumer Research, 31, 209–219.

Sapp, S. G., & Harrod, W. J. (1993). Reliability and validity of a brief version of Levenson's locus of control scale. Psychological Reports, 72, 539–550.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. Journal of Personality and Social Psychology, 67, 1063–1078.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. Science, 185(4157), 1124-1131

Sub-appendix A

Table A1. Experiment 1- Randomization Check

		Experimental conditions				Group differences (p-value)
	Total	Satisfied	Dissatisfied or satisfied	Open	Dissatisfied	
N	*421*	*104*	*99*	*105*	*113*	
Life satisfaction (1-10)[1]	7.2	7.3	7.1	7.2	7.2	0.82
Sat: financial (1-10)[1]	5.6	5.7	5.6	5.4	5.5	0.72
Sat: health (1-10)[1]	8.3	8.4	8.2	8.3	8.6	0.38
Sat: environment (1-10)[1]	7.4	7.6	7.3	7.4	7.4	0.69
Social trust (0-10)[2]	4.3	4.2	4.2	4.4	4.6	0.47
Materialism (1-5)[3]	2.8	2.9	2.9	2.7	2.8	0.15
Age (18-47)	22	23	22	22	23	0.45
Female (%)	68	68	67	75	63	0.26
Partner (%)	59	63	56	57	59	0.69

Notes: Means/percentages are reported. The final column is calculated using a one-way ANOVA.

[1] "All things considered, how satisfied would you say you are with .. (a) your life in general? (b) your financial situation (c) your health situation (d) the city where you live" (1=not at all satisfied; 10=very satisfied).
[2]"Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" (0=You can't be too careful; 10=Most people can be trusted).
[3] 9-item version of the Material Values Scale (Richins 2004).

Table A2. Experiment 2- Randomization Check

	Total	Experimental conditions: Happy	Happy or unhappy	Unhappy or happy	Open	Unhappy	Group differences (p-value)
N	*1,611*	*286*	*360*	*317*	*347*	*301*	
Life satisfaction (1-10)[1]	7.8	7.8	7.7	8.0	7.7	7.7	0.06
Mood (1-10)[2]	7.6	7.5	7.6	7.7	7.6	7.5	0.47
Age (18-92)	56	54	57	56	55	56	0.28
Female (%)	52	52	51	52	52	51	0.99
Partner (%)	71	70	70	74	68	71	0.54
Children (%)	32	34	27	33	32	33	0.26
Tertiary education (%)	39	44	36	38	40	38	0.41
Income > €2600 (%)	48	45	49	46	50	51	0.45
Urban (%)[3]	40	36	41	39	41	40	0.69
Optimism (1-7)[4]	3.4	3.4	3.4	3.4	3.4	3.4	0.84
Materialism (1-7)[5]	2.7	2.7	2.7	2.7	2.7	2.7	0.82
Locus of control (1-7)[6]	3.1	3.1	3.1	3.2	3.1	3.1	0.34

Notes: Means/percentages are reported. The final column is calculated using a one-way ANOVA. The only significant difference (p<0.05) is higher life satisfaction in the *unhappy or happy* condition. The unequal sample size in experiment 2 can be mainly attributed to non-response given that randomization was based on all panel members rather than actual respondents. We used all socio-demographic information provided by CentERdata and all personality variables from the questionnaire. At the beginning of the questionnaire, all respondents answered a non-experimental SWB question about life satisfaction and their mood.

[1] "All things considered, how satisfied are you with your life as a whole?" (1=very dissatisfied; 10=very satisfied)
[2] "How happy do you feel today?" (1=very unhappy; 10=very happy)
[3] Urban=1 if > 1500 residents per km^2
[4] 6-item Revised Life Orientation Test (Scheier et al. 1994)
[5] 9-item version of the Material Values Scale (Richins 2004)
[6] 9-item version of the Levenson IPC scale (Sapp and Harrod 1993)

Table A3. Experiment 3: Life Satisfaction- Randomization Check

		Experimental conditions								
	Total	Satisfied	Dissatisfied or satisfied	Open	Dissatisfied	Satisfied (overall)	Satisfied (short)	Satisfied (overall& unipolar)	Satisfied (short & totally)	Group differences (p-value)
N	*1,043*	*126*	*130*	*136*	*129*	*135*	*128*	*128*	*131*	
SWLS (1-7)[1]	4.9	4.7	4.9	4.8	4.9	4.9	5.0	4.8	4.7	0.50
Cantril ladder (0-10)[2]	6.9	6.8	6.9	6.8	6.8	7.0	7.0	7.0	7.0	0.79
Affect balance (-10; 10)[3]	2.7	2.4	2.8	3.3	2.5	3.0	2.7	2.3	2.6	0.20
Extraversion (1-7)[4]	4.3	4.3	4.5	4.4	4.2	4.4	4.3	4.3	4.3	0.85
Agreeableness (1-7)[4]	4.2	4.3	4.1	4.2	4.1	4.2	4.3	4.0	4.2	0.14
Conscientiousness (1-7)[4]	5.2	5.2	5.2	5.2	5.3	5.0	5.2	5.2	5.1	0.62
Emotional stability (1-7)[4]	4.4	4.5	4.6	4.2	4.3	4.4	4.4	4.4	4.4	0.54
Openness (1-7)[4]	5.2	5.2	5.2	5.3	5.3	5.1	5.2	5.0	5.2	0.72
Optimism (1-7)[5]	4.7	4.7	4.7	4.6	4.7	4.6	4.7	4.5	4.7	0.81
Materialism (1-7)[6]	2.9	2.8	2.9	2.9	2.9	2.8	2.8	2.9	2.8	0.97

Social trust (0-10)[7]	5.8	5.9	6.1	5.9	5.7	5.9	5.9	5.7	5.7	0.86
Health (1-5)[8]	3.9	3.8	3.9	3.9	3.9	3.9	3.9	4.0	3.9	0.79
Social ladder (0-10)[9]	6.9	6.9	6.6	6.9	6.9	7.0	6.9	7.0	6.9	0.70
Worthwhile (0-10)[10]	7.2	7.1	7.2	7.3	7.3	7.2	7.2	7.2	7.0	0.96
Sat: financial (0-10)[11]	6.4	6.3	6.2	6.6	6.2	6.6	6.6	6.2	6.4	0.57
Sat: health (0-10) [11]	7.4	7.2	7.3	7.6	7.2	7.5	7.5	7.4	7.4	0.72
Sat: achievements (0-10) [11]	6.8	6.6	6.9	6.9	6.9	6.7	7.0	6.9	6.6	0.46
Sat: family (0-10) [11]	7.1	6.8	7.2	7.5	7.3	7.1	7.4	7.0	6.9	0.11
Sat: work/study (0-10) [11]	6.9	6.6	7.0	6.9	6.9	6.8	6.9	7.1	6.6	0.42
Sat: safety (0-10) [11]	8.1	8.1	8.1	8.0	8.2	8.2	8.0	8.0	8.0	0.98
Sat: environment (0-10) [11]	7.6	7.5	7.6	7.7	7.6	7.6	7.6	7.6	7.5	0.98
Age (17-88)	24	25	24	25	24	24	23	23	24	0.66
Female (%)	57	57	50	64	58	54	65	49	63	0.10
Partner (%)	40	36	34	47	37	36	54	41	39	0.04

Ethnicity: Dutch (%)	25	21	25	18	19	28	27	34	24	0.07
Lives in Netherlands (%)	46	48	38	48	45	45	45	59	38	0.03
Mother tongue: English (%)	18	17	19	13	22	17	21	16	21	0.66
Source: MOOC (%)[12]	19	20	19	20	19	20	16	19	19	0.99

Notes: Means/percentages are reported. The final column is calculated using a one-way ANOVA.

[1] 5-item satisfaction with life scale (SWLS) of Diener et al. (1985)

[2] "Please imagine a ladder with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. If the top step is 10 and the bottom step is 0, on which step of the ladder do you feel you personally stand at the present time?" (Cantril 1965)

[3] 10-item affect measure as recommended by the OECD guidelines (OECD 2013)

[4] 10-item personality inventory (Gosling et al. 2003)

[5] 6-item Revised Life Orientation Test (Scheier et al. 1994)

[6] 6-item version of the Material Values Scale (Richins 2004)

[7] "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?" (0=You can't be too careful; 10=Most people can be trusted)

[8] "How would you describe your current health?" (0=bad; 10=very good)

[9] "Suppose we say that the top of the ladder represents the top of our society and the bottom of the ladder represents the bottom of our society. If the top step is 10 and the bottom step is 0, where would you place yourself on this scale nowadays?". This measure is taken from ESS 2012

[10] "Overall, to what extent do you feel the things you do in your life are worthwhile?" (0= Not at all worthwhile; 10= Very worthwhile). This is the recommended eudaimonic well-being measure in the OECD guidelines (OECD 2013)

[11] "How satisfied are you with your … (a) financial situation (b) health (c) achievements in life (d) personal relationships (e)work or study (f) feeling of safety (g) local environment" (0=not at all satisfied; 10=completely satisfied)

[12] MOOC=1 if the questionnaire was completed as part of the Massive Open Online Course (MOOC) on critical thinking; MOOC=0 if the questionnaire was completed as part of the Minor Economics of well-being.

Table A4. Experiment 3: Happiness- Randomization Check

		Experimental conditions								
	Total	Happy	Happy or unhappy	Unhappy	Happy (unipolar)	Happy (short; unipolar)	Open (verbal; short)	Open (verbal)	Open (verbal; bipolar)	Group differences (p-value)
$N_{happiness}$	*990*	*132*	*131*	*119*	*113*	*113*	*121*	*124*	*137*	
SWLS (1-7)	4.9	4.9	4.8	4.9	4.9	4.8	4.8	5.0	4.8	0.85
Cantril ladder (0-10)	6.9	7.0	6.9	6.9	7.0	6.8	6.9	6.9	6.9	0.90
Affect balance (-10; 10)	2.7	2.7	2.6	2.9	2.6	2.1	3.0	2.9	2.8	0.33
Extraversion (1-7)	4.3	4.0	4.1	4.4	4.5	4.2	4.6	4.4	4.5	0.00
Agreeableness (1-7)	4.2	4.4	4.3	4.1	4.0	4.3	4.2	4.1	4.2	0.09
Conscientiousness (1-7)	5.2	5.1	5.2	5.2	5.0	5.2	5.1	5.4	5.2	0.45
Emotional stability (1-7)	4.4	4.3	4.3	4.5	4.5	4.3	4.5	4.5	4.4	0.73
Openness (1-7)	5.2	5.2	5.1	5.3	5.2	5.2	5.2	5.2	5.2	0.82
Optimism (1-7)	4.7	4.7	4.6	4.6	4.7	4.7	4.7	4.6	4.6	0.94
Materialism (1-7)	2.9	2.9	2.8	3.0	2.7	3.0	2.8	2.8	2.9	0.17
Social trust (0-10)	5.8	5.7	6.0	5.8	5.9	5.6	5.9	5.7	6.0	0.68
Health (1-5)	3.9	3.9	3.7	3.8	4.0	3.9	4.1	4.0	3.8	0.04
Social ladder (0-10)	6.9	6.6	7.0	7.1	7.0	6.9	6.8	6.9	6.8	0.26
Worthwhile (0-10)	7.2	7.3	7.1	7.2	7.3	7.5	6.8	7.0	7.2	0.06
Sat: financial (0-10)	6.4	6.4	6.9	6.2	6.2	6.5	6.3	6.4	6.3	0.23
Sat: health (0-10)	7.4	7.3	7.0	7.6	7.5	7.6	7.4	7.5	7.3	0.31

Sat: achievements (0-10)	6.8	7.2	6.8	6.9	6.6	6.7	6.7	6.7	6.9	0.24
Sat: family (0-10)	7.1	7.3	7.2	7.1	7.2	7.0	7.2	7.1	7.1	0.99
Sat: work/study (0-10)	6.9	7.2	6.8	7.0	6.7	6.9	6.7	6.6	6.9	0.28
Sat: safety (0-10)	8.1	8.2	8.1	8.2	7.9	8.2	8.2	8.1	7.8	0.61
Sat: environment (0-10)	7.6	7.6	7.5	7.6	7.3	7.7	7.6	7.7	7.6	0.77
Age (17-88)	24	24	24	24	24	23	25	23	24	0.38
Female (%)	58	51	54	65	59	59	57	59	57	0.58
Partner (%)	41	44	44	40	28	36	45	45	40	0.12
Ethnicity: Dutch (%)	26	29	24	24	29	22	26	24	31	0.72
Lives in Netherlands (%)	48	49	46	48	45	48	54	54	45	0.70
Mother tongue: English (%)	19	15	21	20	13	24	24	15	18	0.16
Source: MOOC (%)	19	15	20	18	23	14	25	23	18	0.35

Table A5. Experiment 4- Randomization Check

		Experimental conditions								
	Total	Satisfied	Dissatisfied or satisfied	Open	Dissatisfied	Satisfied (overall)	Satisfied (short)	Satisfied (overall& unipolar)	Satisfied (short & totally)	Group differences (p-value)
N	*4,669*	*615*	*594*	*624*	*540*	*569*	*569*	*537*	*621*	
Happiness (0-10)[1]	6.0	6.0	6.0	6.0	6.0	6.0	5.9	6.0	6.0	0.98
Ideal happiness (0-10)[2]	6.5	6.4	6.5	6.5	6.5	6.5	6.5	6.5	6.5	1.00
Future happiness (-5; 5)[3]	0.1	0.1	0.1	0.3	0.2	0.1	0.0	0.1	0.2	0.21
Health (1-5)[4]	3.6	3.7	3.5	3.6	3.7	3.6	3.5	3.6	3.6	0.10
Sat: job (0-10)[5]	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5	2.6	0.29
Sat: family (0-10)[5]	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	0.59
Sat: partner (0-10)[5]	3.0	2.9	3.0	3.0	3.0	3.0	2.9	2.9	3.0	0.55
Sat: friends (0-10)[5]	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	0.85
Sat: environment (0-10)[5]	2.8	2.9	2.9	2.8	2.8	2.9	2.8	2.8	2.8	0.53

Sat: community (0-10)[5]	2.7	2.8	2.7	2.7	2.7	2.7	2.7	2.7	2.7	0.57
Financial struggle (1-6)[6]	3.3	3.3	3.2	3.3	3.3	3.3	3.4	3.3	3.3	0.23
Worry: COVID (1-4)[7]	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	0.58
Worry: health (1-4)[7]	2.6	2.5	2.6	2.5	2.5	2.6	2.6	2.6	2.6	0.03
Worry: stress (1-4)[7]	2.6	2.6	2.6	2.5	2.5	2.6	2.6	2.6	2.6	0.60
Worry: job loss (1-4)[7]	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	0.99
Worry: wage cut (1-4)[7]	2.3	2.3	2.2	2.2	2.2	2.3	2.3	2.3	2.3	0.99
Worry: bankruptcy (1-4)[7]	2.0	2.0	1.9	2.0	2.0	2.0	2.0	2.0	1.9	0.82
Worry: community (1-4)[7]	1.9	1.9	1.9	1.9	2.0	1.9	2.0	1.9	1.9	0.99
Worry: loans (1-4)[7]	1.7	1.7	1.7	1.8	1.7	1.7	1.8	1.7	1.7	0.92

Worry: child's future (1-4)[7]	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.6	2.5	0.95
Worry: retirement (1-4)[7]	2.7	2.7	2.7	2.6	2.7	2.7	2.7	2.7	2.7	0.91
Household income (log)	4.9	4.9	4.9	5.0	4.9	4.8	5.0	5.0	4.8	0.86
Age (18-93)	46	46	47	46	46	47	47	46	46	0.37
Employed (%)	75	74	75	76	75	75	73	76	75	0.89
Highly educated (%)[8]	53	54	55	55	56	52	53	54	48	0.13
Female (%)	46	47	44	45	45	48	44	43	48	0.47
Partner (%)	53	52	52	56	55	50	54	53	53	0.57
Has children (%)	48	45	49	50	48	44	51	48	46	0.20

Notes: Means/percentages are reported. The final column is calculated using a one-way ANOVA.

[1] "How happy are you currently?" (0=very unhappy; 10=very happy)

[2] "How happy do you think people should be?" (0=feeling only unhappiness; 10=feeling only happiness)

[3] "How happy do you think you will be five years from now compared to now" (-5= five points unhappier than now; 5= five points happier than now)

[4] "How is your current health?" (1=not healthy; 5=healthy)

[5] "How satisfied are you with: … (1=very dissatisfied; 4=very satisfied)"

[6] "How challenging do you find it to afford your daily living expenses?" (1=very easy; 6=very difficult)

[7] Are you currently worried about …? (1=not at all; 4=very)

[8] Having completed an undergraduate degree or higher.

Table A6: Question Tone- Means and Dispersion (Unconditional Results)

	Experiment 1: Greece		Experiment 2: Netherlands		Experiment 3: Students		Experiment 4: Japan	
Condition	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Panel A: Life satisfaction								
Satisfied	7.27	1.61			7.11	1.51	5.87	2.18
Dissatisfied or satisfied	7.03	1.76			7.09	1.42	5.71	2.32
Open	6.87	1.77			6.88	1.91	5.83	2.26
Dissatisfied	7.18	1.49			6.97	1.84	5.59	2.34
Panel B: Happiness								
Happy			5.35	0.82	7.24	1.43		
Happy or unhappy			5.30	0.77	7.02	1.41		
Unhappy or happy			5.38	0.81				
Open			5.20	0.79				
Unhappy			5.26	0.92	6.31	2.22		

Note: Unadjusted means and standard deviations (SD) reported.

Table A7. Question Tone- Bivariate Relationships (DV= Life Satisfaction)

	(1) Satisfied	(2) Dissatisfied or satisfied	(3) Open	(4) Dissatisfied	(5) Significant differences
Experiment 1: Greece					
Materialism	-0.31*	-0.08	-0.16	-0.36**	
Social trust	0.16**	0.13*	0.24**	0.09	3>4
Sat: financial	0.27**	0.31**	0.42**	0.30**	
Sat: health	0.16*	0.22**	0.30**	0.33**	
Sat: environment	0.29**	0.33**	0.47**	0.23**	3>4
Female	-0.08	0.18	0.01	0.06	
Partner	-0.05	0.02	-0.16	0.02	
Experiment 3: Students					
Extraversion	0.22*	0.12	0.13	0.14	
Agreeableness	0.00	0.31*	0.12	0.10	
Conscientiousness	0.24	0.41**	0.20	0.16	
Emotional stability	0.26*	0.29**	0.27*	0.45**	

Openness	-0.04	-0.01	0.03	0.25	
Optimism	0.41**	0.57**	0.70**	0.53**	
Materialism	-0.29	0.05	-0.09	-0.27	
Social trust	0.23**	0.12	0.28**	0.14	
Health	0.32*	0.42**	0.64**	0.42	
Social ladder	0.67**	0.28**	0.27	0.24	1>2+3+4
Worthwhile	0.31**	0.34**	0.50**	0.33*	
Sat: financial	0.27**	0.11	0.28**	0.19**	
Sat: health	0.16*	0.22**	0.30**	0.33**	
Sat: achievements	0.43**	0.26**	0.60**	0.37**	3>2
Sat: family	0.26**	0.29**	0.37**	0.24**	
Sat: work/study	0.45**	0.33**	0.60**	0.37**	3>2
Sat: safety	0.27**	0.27**	0.49**	0.13	3>4
Sat: environment	0.27**	0.25**	0.47**	0.15	3>4
Female	-0.40	0.05	-0.38	0.06	
Partner	0.10	0.42	-0.10	-0.04	
Experiment 4: Japan					
Future happiness	0.49**	0.55**	0.51**	0.55**	
Health	0.89**	0.77**	1.00**	0.92**	
Sat: job	1.31**	1.41**	1.29**	1.48**	
Sat: family	1.33**	1.24**	1.33**	1.15**	
Sat: partner	1.03**	0.79**	1.20**	1.00**	3>2
Sat: friends	1.29**	1.05**	1.35**	1.16**	
Sat: environment	1.39**	1.21**	1.32**	1.18**	
Sat: community	1.16**	1.01**	1.12**	1.13**	
Financial struggle	-0.76**	-0.76**	-0.68**	-0.68**	
Worry: COVID	-0.05	-0.22	-0.13	-0.46**	1+3>4
Worry: health	-0.36**	-0.54**	-0.54**	-0.61**	
Worry: stress	-0.72**	-0.88**	-0.80**	-0.84**	
Worry: job loss	-0.70**	-0.66**	-0.54**	-0.81**	
Worry: wage cut	-0.53**	-0.71**	-0.58**	-0.64**	
Worry: bankruptcy	-0.51**	-0.35**	-0.47**	-0.58**	
Worry: community	-0.09	0.18	-0.17	-0.28*	2>4
Worry: loans	-0.32**	-0.24*	-0.26*	-0.35**	
Worry: child's future	-0.09	-0.23	-0.43**	-0.50**	
Worry: retirement	-0.50**	-0.67**	-0.55**	-0.65**	
HH income (log)	0.11**	0.13**	0.18**	0.11*	
Age	-0.10**	-0.14**	-0.18**	-0.13**	
Age squared	0.12**	0.16**	0.21**	0.13**	
Employed	0.27	0.14	0.46	0.02	
Highly educated	0.47**	0.79**	0.55**	0.72**	

Female	0.33	-0.04	0.21	0.31	
Partner	1.22**	1.12**	1.53**	1.31**	
Has children	1.02**	0.99**	1.40**	1.28**	

Note: Each cell in Columns 1-4 represents a separate OLS regression where life satisfaction is regressed on the covariate. Robust standard errors are used. Column 5 reports statistically significant differences (p<0,05) in coefficients between conditions as indicated by a significant interaction coefficient in a regression of life satisfaction on the experimental condition, the covariate, and its interaction. * p<0.05; ** p<0.01. With 4 conditions, there are 6 pairwise comparisons per covariate. The number of pairwise comparisons is 6 pairs multiplied by the number of covariates. The percentage of pairwise comparisons showing significant differences is 5% in experiment 1 (2 out of 42), 6% in experiment 2 (7 out of 120), 6% for life satisfaction in experiment 3, and 2% in experiment 4 (4 out of 162).

Table A8. Question Tone- Bivariate Relationships (DV= happiness)

	(1) Happy	(2) Happy or unhappy	(3) Unhappy or happy	(4) Open	(5) Unhappy	(6) Significant differences
Experiment 2: Netherlands						
Age	-0.01	-0.04*	0.02	-0.00	-0.02	3>2
Age^2/100	0.01	0.04*	-0.01	0.00	0.03	2>3
Female	0.16	0.07	0.01	-0.11	0.06	1>4
Partner	0.36**	0.42**	0.19	0.45**	0.33*	
Children	0.08	0.12	-0.04	0.12	-0.19	2+4>5
Education	-0.04	0.09	0.09	0.05	0.17	
Income > €2600	0.25**	0.37**	0.26**	0.27**	0.11	
Urban	-0.05	-0.00	-0.10	-0.28**	0.08	2+5>4
Optimism	0.62**	0.73**	0.76**	0.65**	0.65**	
Materialism	-0.15*	-0.18**	-0.30**	-0.20**	-0.16**	
Locus of control	0.51**	0.71**	0.81**	0.75**	0.76**	
Experiment 3: Students						
Extraversion	0.32**	0.14			0.25	
Agreeableness	-0.02	0.08			0.05	
Conscientiousness	0.30**	0.36**			0.21	
Emotional stability	0.31**	0.44**			0.51**	
Openness	0.14	0.08			0.14	
Optimism	0.68**	0.62**			0.66**	
Materialism	-0.02	-0.09			-0.50	
Social trust	0.14*	0.20*			0.30**	
Health	0.54**	0.72**			0.24	

Social ladder	0.33*	0.43**	0.42*
Worthwhile	0.50**	0.36**	0.47**
Sat: financial	0.31**	0.24**	0.29**
Sat: health	0.30**	0.40**	0.35**
Sat: achievements	0.49**	0.46**	0.57**
Sat: family	0.33**	0.39**	0.45**
Sat: work/study	0.40**	0.46**	0.51**
Sat: safety	0.31**	0.46**	0.55**
Sat: environment	0.33**	0.45**	0.47**
Female	0.05	-0.10	0.17
Partner	-0.13	0.16	0.14

Note: The same method is used as in Table A7. With 5 conditions, there are 10 pairwise comparisons per covariate. Therefore, the number of pairwise comparisons in experiment 2 is 10 pairs multiplied by the number of covariates. With 3 conditions, there are 3 pairwise comparisons per covariate. Therefore, the number of pairwise comparisons in experiment 3 is 3 pairs multiplied by the number of covariates. The percentage of pairwise comparisons showing significant differences is 6% in experiment 1 (7 out of 110) and 0% in experiment 3 (0 out of 60).

Table A9. Experiment 3- Mechanism Variables

Mechanism	Question	Scale
Life domains considered	Indicate to what extent the following domains have affected your answer to the question [experimental question]? (1) my financial situation (2) my health (3) my achievements in life (4) my family life (5) my work or study (6) my feeling of safety (7) the quality of my local environment	(a) I didn't think about it when answering the question. (b) I thought about it but decided it was not important for choosing my answer. (c) I thought about it and decided it was somewhat important for choosing my answer. (d) I thought about it and decided it was very important for choosing my answer.
Perceived norm	Consider again the question [experimental question]? What would you estimate to be the average score given by people in your country of residence to this question?	Response scale as in experimental condition
Scale interpretation	Consider again the question *[experimental question]?* Only the lowest and highest number on the response scale were labeled. We could also assign labels to the other numbers on the scale. (1) In your opinion, what number on the scale corresponds to being "a bit satisfied" with life? (2) And what number on the scale would correspond to being “a bit dissatisfied” with life?	Response scale as in experimental condition
Question difficulty	Consider again the question *[experimental question]*? How difficult was it for you to understand the question and the response scale?	(1) Easy (2) Neither easy nor difficult (3) Difficult
Response duration	Recorded duration by Qualtrics until response submission of the experimental question	In seconds

Table A10. Other Question and Scale Wording Differences- Means and Dispersion (Unconditional Results)

	Experiment 3		**Experiment 4**	
Condition	*Mean*	*SD*	*Mean*	*SD*
Panel A: Life satisfaction				
Satisfied	7.11	1.51	5.87	2.18
Satisfied (short)	7.17	1.70	5.85	2.26
Satisfied (overall)	7.17	1.39	5.73	2.29
Satisfied (overall; unipolar)	7.01	1.44	5.82	2.22
Satisfied (short; totally)	7.14	1.54	5.83	2.26
Panel B: Happiness				
Happy	7.14	0.89		
Happy (unipolar)	7.20	1.54		
Happy (short; unipolar)	7.27	1.35		
Open (verbal; short)	3.17	0.65		
Open (verbal)	3.15	0.67		
Open (verbal; bipolar)	2.99	0.59		

Note: Unadjusted means and standard deviations (SD) reported. The 4-item happiness scale is reverse scored in all analysis so that higher values reflect higher happiness.

Table A11. Other Wording and Scale Differences- Bivariate Relationships (DV= Life Satisfaction)

	(1) Satisfied	(2) Satisfied (short)	(3) Satisfied (overall)	(4) Satisfied (overall; unipolar)	(5) Satisfied (short; totally)	(6) Significant differences
Experiment 3: students						
Extraversion	0.22*	0.23*	0.26**	0.22*	0.40**	
Agreeableness	0.00	-0.03	0.19	-0.33**	-0.21	4<1+3
Conscientiousness	0.24	0.19	0.24*	0.19	0.31**	
Emotional stability	0.26*	0.51**	0.27**	0.40**	0.38**	
Openness	-0.04	0.31	0.20	0.23	0.15	
Optimism	0.41**	0.69**	0.50**	0.74**	0.74**	
Materialism	-0.29	-0.16	-0.08	-0.08	-0.07	
Social trust	0.23**	0.32**	0.18**	0.21**	0.23**	
Health	0.32*	0.76**	0.48**	0.46*	0.35	
Social ladder	0.67**	0.34*	0.31**	0.49**	0.36**	1>3
Worthwhile	0.31**	0.42**	0.31**	0.53**	0.38**	1<4
Sat: financial	0.27**	0.30**	0.25**	0.23**	0.24**	
Sat: health	0.16*	0.42**	0.30**	0.29**	0.45**	1<2+5
Sat: achievements	0.43**	0.54**	0.35**	0.43**	0.42**	
Sat: family	0.26**	0.50**	0.33**	0.45**	0.40**	1<2+4
Sat: work/study	0.45**	0.53**	0.20*	0.40**	0.29**	2>3
Sat: safety	0.27**	0.34*	0.23*	0.28**	0.37**	
Sat: environment	0.27**	0.43**	0.24*	0.37**	0.46**	
Female	-0.40	0.15	0.09	-0.02	-0.01	
Partner	0.10	0.28	0.04	0.46	0.48	

Experiment 4: Japan						
Future happiness	0.49**	0.64**	0.57**	0.58**	0.50**	2>1+5
Health	0.89**	0.94**	1.00**	0.88**	0.96**	
Sat: job	1.31**	1.48**	1.31**	1.49**	1.54**	
Sat: family	1.33**	1.39**	1.49**	1.29**	1.48**	
Sat: partner	1.03**	1.01**	1.07**	0.91**	1.13**	
Sat: friends	1.29**	1.37**	1.64**	1.35**	1.59**	3>1
Sat: environment	1.39**	1.49**	1.36**	1.39**	1.33**	
Sat: community	1.16**	1.08**	1.19**	1.43**	1.28**	
Financial struggle	-0.76**	-0.65**	-0.68**	-0.94**	-0.78**	4<2+3
Worry: COVID	-0.05	-0.20	-0.17	-0.11	-0.33**	
Worry: health	-0.36**	-0.49**	-0.75**	-0.53**	-0.63**	1>3
Worry: stress	-0.72**	-0.84**	-1.07**	-0.78**	-0.78**	3<1+5
Worry: job loss	-0.70**	-0.72**	-0.71**	-0.89**	-0.61**	
Worry: wage cut	-0.53**	-0.71**	-0.57**	-0.70**	-0.65**	
Worry: bankruptcy	-0.51**	-0.46**	-0.40**	-0.55**	-0.42**	
Worry: community	-0.09	-0.23	-0.29*	-0.14	-0.09	
Worry: loans	-0.32**	-0.29*	-0.45**	-0.36**	-0.39**	
Worry: child's future	-0.09	-0.43**	-0.39**	-0.22	-0.23	
Worry: retirement	-0.50**	-0.51**	-0.68**	-0.57**	-0.54**	
HH income (log)	0.11**	0.12*	0.15**	0.09	0.15**	
Age	-0.10**	-0.13**	-0.11**	-0.15**	-0.11**	
Age squared	0.12**	0.15**	0.13**	0.17**	0.14**	
Employed	0.27	0.18	-0.10	-0.21	0.37	
Highly educated	0.47**	0.46*	0.24	0.37	0.06	
Female	0.33	0.06	0.44*	0.70**	0.15	4>2+5

Partner	1.22**	1.61**	0.98**	1.35**	1.14**	2>3
Has children	1.02**	1.28**	0.90**	1.09**	0.99**	

Note: The same method is used as in Table A7. The percentage of pairwise comparisons showing significant differences is 5% in experiment 3 (9 out of 200) and 4% in experiment 4 (11 out of 270).

Table A12. Other Wording and Scale Differences- Bivariate Relationships (DV= happiness)

	(1) Happy	(2) Happy (unipolar)	(3) Happy (short; unipolar)	(4) Open (verbal; short)	(5) Open (verbal)	(6) Open (verbal; bipolar)	(7) Significant differences
Extraversion	0.32**	0.30**	0.21*	0.14**	0.20**	0.11**	
Agreeableness	-0.02	-0.14	0.05	0.15**	0.03	0.02	
Conscientiousness	0.30**	0.09	0.32**	0.14**	0.07	0.10**	
Emotional stability	0.31**	0.26*	0.35**	0.18**	0.13**	0.17**	
Openness	0.14	0.10	0.20	0.06	0.08	0.13**	
Optimism	0.68**	0.66**	0.54**	0.24**	0.27**	0.27**	
Materialism	-0.02	0.06	0.16	-0.08	-0.08	-0.17*	
Social trust	0.14*	0.32**	0.01	0.07**	0.05	0.08**	2>1+3
Health	0.54**	0.55**	0.44**	0.16*	0.30**	0.22**	
Social ladder	0.33*	0.50**	0.29**	0.15**	0.12**	0.07	
Worthwhile	0.50**	0.45**	0.34**	0.13**	0.17**	0.12**	

Sat: financial	0.31**	0.20**	0.00	0.08**	0.07*	0.06*	3<1+2
Sat: health	0.30**	0.29**	0.27**	0.10**	0.00**	0.11**	
Sat: achievements	0.49**	0.51**	0.30**	0.15**	0.15**	0.11**	3<1+2
Sat: family	0.33**	0.52**	0.38**	0.19**	0.13**	0.11**	2>1; 4>6
Sat: work/study	0.40**	0.46**	0.39**	0.13**	0.12**	0.11**	
Sat: safety	0.31**	0.40**	0.19**	0.12**	0.11**	0.08*	
Sat: environment	0.33**	0.35**	0.19**	0.15**	0.11**	0.12**	
Female	0.05	-0.17	-0.18	0.27*	-0.02	-0.02	
Partner	-0.13	0.24	0.83**	0.12	0.06	0.13	3>1

Note: The same method is used as in Table A7. Data are from experiment 3. Columns 1-3 are not compared to Columns 4-6 because of the different scales. With 3 conditions, there are 3 pairwise comparisons per covariate for Columns 1-3. Similarly, 3 pairwise comparisons per covariate are made for Columns 4-6. Therefore, the number of pairwise comparisons is 6 pairs multiplied by 20 covariates. The percentage of pairwise comparisons showing significant differences is 7,5% (9 out of 120).

Table A13. Other Wording and Scale Differences- Mechanisms

	(1) Satisfied	(2) Satisfied (short)	(3) Satisfied (overall)	(4) Satisfied (overall; unipolar)	(5) Satisfied (short; totally)	(6) Significant differences
Focus: financial (%)[1]	26	27	24	27	29	
Focus: health (%)	43	38	45	43	40	

Focus: achievements (%)	46	41	39	37	45	
Focus: family (%)	48	69	60	62	69	2+4+5>1
Focus: work/study (%)	53	50	37	48	45	3 <1+2
Focus: safety (%)	17	22	23	24	22	
Focus: environment (%)	31	28	24	29	30	
Domain score of focus area[2]	6.8	7.2	7.4	7.2	6.9	3>1
Perceived norm	6.3	6.5	6.5	6.5	6.4	
Scale interpretation: A bit satisfied	5.7	5.5	5.7	5.6	5.7	
Scale interpretation: A bit dissatisfied	4.1	4.0	4.2	4.2	4.1	
Response duration[3]	11.4	9.4	10.0	9.7	8.8	
Question difficulty	1.7	1.5	1.6	1.7	1.6	1>2

[1] The domain percentages reflect the percentage of respondents considering the domain "very important" for choosing their answer.
[2] This score is calculated by a person's average domain satisfaction on domains they considered "very important" for choosing their answer. Hence, a lower score means that domains a respondent was less satisfied with was considered more when answering the life satisfaction question.
[3] Medians are reported for response duration and differences calculated using a Dunn's test to minimize effects of outliers.

Sub-appendix B
Other Variables Included in Larger Database

Table B1. Description of Variables

Name	Varlab
year	Year
LifeSatisfaction	LifeSatisfaction
AgeExact	AgeExact
MaritalStatus	MaritalStatus
AgeEducation	AgeEducation
SexGender	SexGender
LeftRightPlacement	LeftRightPlacement
DemocracySatisfactionCountry	DemocracySatisfactionCountry
DemocracySatisfactionEU	DemocracySatisfactionEU
region_code	
level	Nuts_Level
SECTOR	SECTOR
NAME_HTML	Nuts Name
Pop_civ_lbf	Active Population - Civilian Labour Force
Pop_tot_lbf	Active Population - Total Labour Force
Pop_av_ann	Average annual population
Capital	Capital Stock at constant prices
Civil_empl	Civilian Employment
Comp_cons	Compensation of Employees at constant prices
Comp_current	Compensation of employees at current prices
Comp_hour	Compensation of employees per hour worked
Fixed_capital_cons	Consumption of fixed capital at constant prices
Fixed_capital_current	Consumption of fixed capital at current prices

Taxes	Current taxes on income and wealth
GDP_growthrate	GDP Growth rate %
GDP_current	GDP at current prices
GDP_cap_constant	GDP per capita at constant prices
GDP_cap_current	GDP per capita at current prices
GVA_growth	GVA Growth rate %
GVA_constant	GVA at constant prices
GVA_current	GVA at current prices
GFC_constant	Gross Fixed Capital Formation at constant prices
GFC_current	Gross Fixed Capital Formation at current prices
Hours_worked_employed	Hours Worked (employed persons)
Hours_workerd_employees	Hours Worked (employees)
Hours_worked_capita	Hours worked per capita
Hours_worked_employed	Hours worked per employed person
Household_disposable_inc	Households net disposable income
Net_property_inc	Net property income
Nom_comp_employee	Nominal compensation per employee
Nom_prod_labor_employee	Nominal labour productivity per hour worked
Nom_prod_labor_employed	Nominal labour productivity per person employed
Unit_labor_cost_per	Nominal unit labour cost based on hours worked
Unit_labor_cost_per	Nominal unit labour cost based on persons
population	Population on 1st January
Comp_employee	Real compensation per employee
Real_prod_labor_hour	Real labour productivity per hour worked
Real_prod_labor_employed	Real labour productivity per person employed
Employment	Total Employment

Wage_salary	Wage and salary earners
region_name	Name of the region in local language
nuts0	Nuts Level 0
nuts1	Nuts Level 1
nuts2	Nuts Level 2
cname	Name of the country in English
cri_contr	Number of awarded contracts above 130,000 EUR
cri_cvalue	Final value of awarded tenders of over 130,000 EUR
cri_singleb	Share of contracts with only one bid in total
cri_nocall	Share of contracts with no published call for tender red flag
cri_nonopen	Share of contracts with non-open procedure red flag
cri_taxhav	Share of contracts with tax haven red flag
eqi_score	EQI Index Score
eqi_zquality	Quality pillar, country centered and z-score standardized
eqi_zimpartiality	Impartiality pillar, country centered and z-score standardized
eqi_zcorruption	Corruption pillar, country centered and z-score standardized
eqi_zcorruptper	Corruption perceptions index (corruption sub-pillar) z-score stand. (2017 only)
eqi_zcorruptexp	Corruption experiences index (corruption sub-pillar) z-score stand. (2017 only)
eqi_norm_eqi	EQI index, min-max (0-100) standardized
eqi_norm_qual	Quality pillar, country centered and min-max (0-100) standardized

eqi_norm_impart	Impartiality pillar, country centered and min-max (0-100) standardized
eqi_norm_corrupt	Corruption pillar, country centered and min-max (0-100) standardized
eqi_norm_corruptper	Corruption perceptions index (corruption sub-pillar) min-max (0-100)(2017)
eqi_norm_corruptexp	Corruption experiences index (corruption sub-pillar) min-max (0-100) (2017)
eu_cri_bur	Reported number of cases of burglary of private premises
eu_cri_inthom	Reported number of cases of intentional homicide
eu_cri_rob	Reported number of cases of robbery
eu_agemoth	Proportion of live births outside marriage
eu_agemoth1	Total fertility rate
eu_nmarpct	Mean age of women at childbirth
eu_totferrt	Mean age of women at birth of first child
eu_death_totalf	Number of deaths of females, all ages
eu_death_totalm	Number of deaths of males, all ages
eu_death_totalt	Number of deaths, total all ages
eu_death_y1f	Number of deaths of females, at 1 year old
eu_death_y1m	Number of deaths of males, at 1 year old
eu_death_y1t	Number of deaths, total at 1 year old
eu_death_y20f	Number of deaths of females, at 20 years old
eu_death_y20m	Number of deaths of males, at 20 years old
eu_death_y20t	Number of deaths, total at 20 years old
eu_death_y50f	Number of deaths of females, at 50 years old
eu_death_y50m	Number of deaths of males, at 50 years old
eu_death_y50t	Number of deaths, total at 50 years old
eu_death_y70f	Number of deaths of females, at 70 years old

eu_death_y70m	Number of deaths of males, at 70 years old
eu_death_y70t	Number of deaths, total at 70 years old
eu_d2jan_f	Population at 1st January, female
eu_d2jan_m	Population at 1st January, male
eu_d2jan_t	Population at 1st January, total
eu_d3area_lat	Area of a region, land area total, sq km
eu_d3area_t	Area of a region, total, sq km
eu_per_km2	Population density, average population per square km
eu_frate_total	Fertility rate, total
eu_frate_y15	Fertility rate, at age 15
eu_frate_y30	Fertility rate, at age 30
eu_frate_y35	Fertility rate, at age 35
eu_mlifexp_f	Life expectancy in years at 1 year old, female
eu_mlifexp_m	Life expectancy in years at 1 year old, male
eu_mlifexp_t	Life expectancy in years at 1 year old, total
eu_edatt_ed02_y2564f	Educational attainment for ages 25 to 64, primary education, female
eu_edatt_ed02_y2564m	Educational attainment for ages 25 to 64, primary education, male
eu_edatt_ed02_y2564t	Educational attainment for ages 25 to 64, primary education, total
eu_edatt_ed34_y2564f	Educational attainment for ages 25 to 64, secondary education, female
eu_edatt_ed34_y2564m	Educational attainment for ages 25 to 64, secondary education, male
eu_edatt_ed34_y2564t	Educational attainment for ages 25 to 64, secondary education, total
eu_edatt_ed58_y2564f	Educational attainment for ages 25 to 64, tertiary education, female

eu_edatt_ed58_y2564m	Educational attainment for ages 25 to 64, tertiary education, male
eu_edatt_ed58_y2564t	Educational attainment for ages 25 to 64, tertiary education, total
eu_edatt_ed02_y3034f	Educational attainment for ages 30 to 34, primary education, female
eu_edatt_ed02_y3034m	Educational attainment for ages 30 to 34, primary education, male
eu_edatt_ed02_y3034t	Educational attainment for ages 30 to 34, primary education, total
eu_edatt_ed34_y3034f	Educational attainment for ages 30 to 34, secondary education, female
eu_edatt_ed34_y3034m	Educational attainment for ages 30 to 34, secondary education, male
eu_edatt_ed34_y3034t	Educational attainment for ages 30 to 34, secondary education, total
eu_edatt_ed58_y3034f	Educational attainment for ages 30 to 34, tertiary education, female
eu_edatt_ed58_y3034m	Educational attainment for ages 30 to 34, tertiary education, male
eu_edatt_ed58_y3034t	Educational attainment for ages 30 to 34, tertiary education, total
eu_eduleave_f	Early leavers from education and training as a percentage, females
eu_eduleave_m	Early leavers from education and training as a percentage, males
eu_eduleave_t	Early leavers from education and training as a percentage, total
eu_neet_y1524f	15-24 year old neither in employment nor in education as percentage, female
eu_neet_y1524m	15-24 year old neither in employment nor in education as percentage, male

eu_neet_y1524t	15-24 year old neither in employment nor in education as percentage, total
eu_empl_durtotal	Employment rate for people between 15-34 years, total duration since education
eu_empl_dury_gt3	Employment rate for people between 15-34 years, over 3 years since education
eu_empl_dury13	Employment rate for people between 15-34 years, 1 to 3 years since education
eu_empl_edled02	Employment rate for people between 15-34 years, education levels 0-2
eu_empl_edled34	Employment rate for people between 15-34 years, education levels 3-4
eu_empl_edled58	Employment rate for people between 15-34 years, education levels 5-8
eu_empl_edltotal	Employment rate for people between 15-34 years, all education levels
eu_epred12	Participation rate in Primary and lower secondary education
eu_epred58	Participation rate in Tertiary education
eu_env_wasdsp_i	Municipal waste disposal - incineration in thousand tonnes
eu_env_wasgen	Municipal waste generated in thousand tonnes
eu_env_wasrcv_e	Municipal waste recovery - energy recovery in thousand tonnes
eu_env_wasrcy_c_d	Municipal waste recycling in thousand tonnes
eu_hea_cs_f	Number of deaths by circulatory system diseases, female
eu_hea_cs_m	Number of deaths by circulatory system diseases, male
eu_hea_cs_t	Number of deaths by circulatory system diseases, total
eu_hea_hiv_f	Number of deaths by HIV, female

eu_hea_hiv_m	Number of deaths by HIV, male
eu_hea_hiv_t	Number of deaths by HIV, total
eu_hea_ipd_f	Number of deaths by infectious and parasitic diseases, female
eu_hea_ipd_m	Number of deaths by infectious and parasitic diseases, male
eu_hea_ipd_t	Number of deaths by infectious and parasitic diseases, total
eu_hea_np_f	Number of deaths by malignant neoplasms, female
eu_hea_np_m	Number of deaths by malignant neoplasms, male
eu_hea_np_t	Number of deaths by malignant neoplasms, total
eu_hea_ns_f	Number of deaths by nervous system diseases, female
eu_hea_ns_m	Number of deaths by nervous system diseases, male
eu_hea_ns_t	Number of deaths by nervous system diseases, total
eu_hea_pr_f	Number of deaths by pregnancy, childbirth and puerperium
eu_hea_sh_f	Number of deaths by self-harm, female
eu_hea_sh_m	Number of deaths by self-harm, male
eu_hea_sh_t	Number of deaths by self-harm, total
eu_hea_tox_f	Number of deaths by drug dependence, female
eu_hea_tox_m	Number of deaths by drug dependence, male
eu_hea_tox_t	Number of deaths by drug dependence, total
eu_hea_bed	Available beds in hospitals (HP.1) per hundred thousand inhabitants

eu_hea_bedcur	Curative care beds in hospitals (HP.1) per hundred thousand inhabitants
eu_hea_bedlt	Long-term care beds in hospitals (HP.1) per hundred thousand inhabitants
eu_hea_bedoth	Other beds in hospitals (HP.1) per hundred thousand inhabitants
eu_hea_bedpsy	Psychiatric care beds in hospitals (HP.1) per hundred thousand inhabitants
eu_hea_bedreh	Rehabilitative care beds in hospitals (HP.1) per hundred thousand inhabitants
eu_hea_dent	Dentists per hundred thousand inhabitants
eu_hea_mdoc	Medical doctors per hundred thousand inhabitants
eu_hea_nurs	Nurses and midwives per hundred thousand inhabitants
eu_hea_pharm	Pharmacists per hundred thousand inhabitants
eu_hea_phys	Physiotherapists per hundred thousand inhabitants
eu_emtk_ab_f	Employment in agriculture, fishing and mining, % of tot. employment, female
eu_emtk_ab_m	Employment in agriculture, fishing and mining, % of tot. employment, male
eu_emtk_ab_t	Employment in agriculture, fishing and mining, % of tot. employment, total
eu_emtk_c_f	Employment in manufacturing, % of tot. employment, female
eu_emtk_c_m	Employment in manufacturing, % of tot. employment, male
eu_emtk_c_t	Employment in manufacturing, % of tot. employment, total
eu_emtk_chtc_f	Employment in high-technology manufacturing, % of tot. employment, female

eu_emtk_chtc_m	Employment in high-technology manufacturing, % of tot. employment, male
eu_emtk_chtc_t	Employment in high-technology manufacturing, % of tot. employment, total
eu_emtk_df_f	Employment in electricity, gas and water supply, % of tot. employment, female
eu_emtk_df_m	Employment in electricity, gas and water supply, % of tot. employment, male
eu_emtk_df_t	Employment in electricity, gas and water supply, % of tot. employment, total
eu_emtk_gu_f	Employment in services, % of tot. employment, female
eu_emtk_gu_m	Employment in services, % of tot. employment, male
eu_emtk_gu_t	Employment in services, % of tot. employment, total
eu_emtk_htc_f	Employment in high-technology sectors, % of tot. employment, female
eu_emtk_htc_m	Employment in high-technology sectors, % of tot. employment, male
eu_emtk_htc_t	Employment in high-technology sectors, % of tot. employment, total
eu_emtk_j_f	Employment in information and communication, % of tot. employment, female
eu_emtk_j_m	Employment in information and communication, % of tot. employment, male
eu_emtk_j_t	Employment in information and communication, % of tot. employment, total
eu_emtk_k_f	Employment in financial and insurance activities, % of tot. employment, female
eu_emtk_k_m	Employment in financial and insurance activities of tot. employment, male

eu_emtk_k_t	Employment in financial and insurance activities, % of tot. employment, total
eu_emtk_kis_f	Employment in knowledge-intensive services, % of tot. employment, female
eu_emtk_kis_m	Employment in knowledge-intensive services, % of tot. employment, male
eu_emtk_kis_t	Employment in knowledge-intensive services, % of tot. employment, total
eu_emtk_kl_f	Employment in real estate activities, % of tot. employment, female
eu_emtk_kl_m	Employment in real estate activities, % of tot. employment, male
eu_emtk_kl_t	Employment in real estate activities, % of tot. employment, total
eu_emtk_m_f	Employment in scientific and technical activities, % of tot. employment, female
eu_emtk_m_m	Employment in scientific and technical activities, % of tot. employment, male
eu_emtk_m_t	Employment in scientific and technical activities, % of tot. employment, total
eu_emtk_n_f	Employment in admin. and support activities, % of tot. employment, female
eu_emtk_n_m	Employment in admin. and support activities, % of tot. employment, male
eu_emtk_n_t	Employment in admin. and support activities, % of tot. employment, total
eu_emtk_ou_f	Employment in extraterritorial org. and bodies, % of tot. employment, female
eu_emtk_ou_m	Employment in extraterritorial org. and bodies, % of tot. employment, male
eu_emtk_ou_t	Employment in extraterritorial org. and bodies, % of tot. employment, total

eu_emtk_p_f	Employment in education, % of tot. employment, female
eu_emtk_p_m	Employment in education, % of tot. employment, male
eu_emtk_p_t	Employment in education, % of tot. employment, total
eu_emtk_q_f	Employment in health and social work activities, % of tot. employment, female
eu_emtk_q_m	Employment in health and social work activities, % of tot. employment, male
eu_emtk_q_t	Employment in health and social work activities, % of tot. employment, total
eu_emtk_r_f	Employment in arts, entertainment and recreation, % of tot. employment, female
eu_emtk_r_m	Employment in arts, entertainment and recreation, % of tot. employment, male
eu_emtk_r_t	Employment in arts, entertainment and recreation, % of tot. employment, total
eu_emtk_s_f	Employment in other service activities, % of tot. employment, female
eu_emtk_s_m	Employment in other service activities, % of tot. employment, male
eu_emtk_s_t	Employment in other service activities, % of tot. employment, total
eu_povrisk_pc	At-risk-of-poverty rate by NUTS regions, percentage
eu_lwoin_pc	People (0 to 59 years) in households with low work intensity, as %
eu_lwoin_pc_y_lt60	People (0 to 59 years) in households with low work intensity, % of total pop.
eu_matdep_pc	Severe material deprivation rate by NUTS regions, percentage

eu_povr_pc	People at risk of poverty or social exclusion by NUTS regions, percentage
eu_igs_b3_12	Last online purchase: between 3 and 12 months ago, percentage
eu_igs_bfeu	Online purchases: from sellers from other EU countries, percentage
eu_igs_bhols	Online purchases: travel and holiday accommodation, percentage
eu_igs_blt12	Last online purchase: in the 12 months, percentage
eu_igs_bumt12x	Last online purchase: more than a year ago or never, percentage
eu_igs_buy3	Last online purchase: in the last 3 months, percentage
eu_is_bacc	Percentage of households with broadband internet access
eu_iu_never	Percentage of individuals who have never used a computer
eu_iu_govform	Percentage of individuals using internet to interact with public authorities
eu_iu_govint	Percentage of individuals using internet to submit forms to authorities
eu_is_iacc	Percentage of households with internet access
eu_iu_ohw	Individuals who accessed internet away from home or work, %
eu_iu_ohw3	Individuals who accessed internet away from home or work in the last 3 months, %
eu_iu_iday	Frequency of internet access: daily
eu_iu_ilt12	Last internet use: in the last 12 months
eu_iu_iu3	Last internet use: in last 3 months
eu_iu_iubk	Internet use: Internet banking
eu_iu_iucpp	Internet use: civic or political participation

eu_iu_iuse	Frequency of internet access: once a week (including every day)
eu_iu_iusell	Internet use: selling goods or services
eu_iu_iusnet	Internet use: participating in social networks
eu_iu_iux	Internet use: never
eu_emp_ft_f	Full-time employment, female, in thousands
eu_emp_ft_m	Full-time employment, male, in thousands
eu_emp_ft_t	Full-time employment, total, in thousands
eu_emp_pt_f	Part-time employment, female, in thousands
eu_emp_pt_m	Part-time employment, male, in thousands
eu_emp_pt_t	Part-time employment, total, in thousands
eu_emp_1524f	Employment rate for 15-24 years old, female
eu_emp_1524m	Employment rate for 15-24 years old, male
eu_emp_1524t	Employment rate for 15-24 years old, total
eu_emp_2064f	Employment rate for 20-64 years old, female
eu_emp_2064m	Employment rate for 20-64 years old, male
eu_emp_2064t	Employment rate for 20-64 years old, total
eu_emp_2534f	Employment rate for 25-34 years old, female
eu_emp_2534m	Employment rate for 25-34 years old, male
eu_emp_2534t	Employment rate for 25-34 years old, total
eu_emp_ge25f	Employment rate for +25 years, female
eu_emp_ge25m	Employment rate for +25 years, male
eu_emp_ge25t	Employment rate for +25 years, total
eu_emp_ge65f	Employment rate for +65 years, female
eu_emp_ge65m	Employment rate for +65 years, male
eu_emp_ge65t	Employment rate for +65 years, total
eu_emp_a	Employment in agriculture, forestry and fishing, in thousands

eu_emp_be	Employment in industry (except construction), in thousands
eu_emp_f	Employment in construction, in thousands
eu_emp_gi	Employment in wholesale and retail trade, and service activities, in thousands
eu_emp_j	Employment in information and communication, in thousands
eu_emp_k	Employment in financial and insurance activities, in thousands
eu_emp_l	Employment in real estate activities, in thousands
eu_emp_m_n	Employment in professional, scientific and technical activities, in thousands
eu_emp_oq	Employment in public admin., defence, education and health, in thousands
eu_emp_ru	Employment in arts, entertainment and recreation, in thousands
eu_emp_total	Employment in all NACE activities, in thousands
eu_ltu_pc_act	Long-term unemployment as percentage of active population
eu_ltu_pc_une	Long-term unemployment as percentage of unemployment
eu_ltu_ths	Long-term unemployment in thousands
eu_unemp_1524f	Unemployment rate for 15-24 years old, female
eu_unemp_1524m	Unemployment rate for 15-24 years old, male
eu_unemp_1524t	Unemployment rate for 15-24 years old, total
eu_unemp_1574f	Unemployment rate for 15-74 years old, female
eu_unemp_1574m	Unemployment rate for 15-74 years old, male
eu_unemp_1574t	Unemployment rate for 15-74 years old, total
eu_unemp_2064f	Unemployment rate for 20-64 years old, female
eu_unemp_2064m	Unemployment rate for 20-64 years old, male

eu_unemp_2064t	Unemployment rate for 20-64 years old, total
eu_unemp_ge15f	Unemployment rate for + 15 years, female
eu_unemp_ge15m	Unemployment rate for + 15 years, male
eu_unemp_ge15t	Unemployment rate for + 15 years, total
eu_unemp_ge25f	Unemployment rate for + 25 years, female
eu_unemp_ge25m	Unemployment rate for + 25 years, male
eu_unemp_ge25t	Unemployment rate for + 25 years, total
eu_b5n_eur_hab	Income of households (Balance) in euro per inhabitant
eu_b5n_mio_eur	Income of households (Balance) in million euro
eu_b5n_mio_nac	Income of households (Balance) in million national currency
eu_b5n_mio_pps	Income of households (Balance) in million PPS
eu_b6n_eur_hab	Income of households (Disposable income) in euro per inhabitant
eu_b6n_mio_eur	Income of households (Disposable income) in million euro
eu_b6n_mio_nac	Income of households (Disposable income) in million national currency
eu_b6n_mio_pps	Income of households (Disposable income) in million PPS
eu_b7n_mio_eur	Income of households (Adjusted disposable income) in million euro
eu_b7n_mio_nac	Income of households (Adjusted disposable income) in million national currency
eu_eng_cdd	Number of cooling degree days
eu_eng_hdd	Number of heating degree days
eu_rdexp_bes	Business enterprise sector intramural expenditure in R&D, euro per inhabitant
eu_rdexp_gov	Government sector intramural expenditure in R&D, euro per inhabitant

eu_rdexp_hes	Higher education sector intramural expenditure in R&D, euro per inhabitant
eu_rdexp_pnp	Private non-profit sector intramural expenditure in R&D, euro per inhabitant
eu_rdexp_total	All sectors intramural expenditure in R&D, euro per inhabitant
eu_prd_bes_f	Total R&D employees in business enterprise sector, female, full-time equivalent
eu_prd_bes_t	Total R&D employees in business enterprise sector, total, full-time equivalent
eu_prd_gov_f	Total R&D employees in government sector, female, full-time equivalent
eu_prd_gov_t	Total R&D employees in government sector, total, full-time equivalent
eu_prd_hes_f	Total R&D employees in higher education sector, female, full-time equivalent
eu_prd_hes_t	Total R&D employees in higher education sector, total, full-time equivalent
eu_prd_pnp_f	Total R&D employees in private non-profit sector, female, full-time equivalent
eu_prd_pnp_t	Total R&D employees in private non-profit sector, total, full-time equivalent
eu_prd_total_f	Total R&D employees in all sectors, female, full-time equivalent
eu_prd_total_t	Total R&D employees in all sectors, total, full-time equivalent
eu_mio_eur	Regional gross domestic product by NUTS 2 regions, million EUR
eu_gdp_mio_pps	Regional gross domestic product (million PPS) by NUTS 2 regions
eu_gdp_pps_hab	Regional gross domestic product (PPS per inhabitant) by NUTS 2 regions

eu_gdp_pps_hab_eu27_2020	Regional gross domestic product. PPS per inhabitant in percentage of EU27
eu_dinc_pps_hab	Disposable income of private households by NUTS 2 regions
eu_pinc_pps_hab	Primary income of private households by NUTS 2 regions
eu_rgva_pch_pre	Real growth rate of regional gross value added (GVA) at basic prices
eu_cnmigratrt	Crude rate of net migration plus statistical adjustment
eu_growrt	Crude rate of total population change
eu_natgrowrt	Crude rate of natural change of population
eu_tour_nstour_bedpl	Number of bed-places in hotels, camping places and other
eu_tour_nstour_estbl	Number of establishments in hotels, camping places and other
eu_tour_bedpl	Net occupancy rate of bed-places in hotels and similar
eu_tour_bedrm	Net occupancy rate of bedrooms in hotels and similar
eu_tour_nscamp	Number of nights spent at camping grounds, recreational vehicle and trailer park
eu_tour_nshotel	Number of nights spent at hotels and similar accommodation
eu_tour_nssa	Number of nights spent at holiday and other short-stay accommodation
eu_tour_nstour	Number of nights spent at tourist accommodations
eu_rac_inj	Injured victims in road accidents, per million inhabitants
eu_rac_kil	Killed victims in road accidents, per million inhabitants

eu_atf_frm_ld	Air transport of freight and mail loaded, in thousand tonnes
eu_atf_frm_ld_nld	Air transport of freight and mail loaded and unloaded, in thousand tonnes
eu_atf_frm_nld	Air transport of freight and mail loaded, in thousand tonnes
eu_mtp_pas_crd	Passengers carried by air transport, in thousand passengers
eu_mtp_pas_crd_arr	Passengers carried by air transport, in thousand passengers
eu_mtp_pas_crd_dep	Passengers carried by air transport (arrival), in thousand passengers
eu_mtf_fr_ld	Maritime transport of freight and mail loaded, in thousand tonnes
eu_mtf_fr_ld_nld	Maritime transport of freight and mail loaded, in thousand tonnes
eu_mtf_fr_nld	Maritime transport of freight and mail loaded and unloaded, in thousand tonnes
eu_mtp_pas	Maritime transport of passengers embarked and disembarked, in thousand passenger
eu_mtp_pas_demb	Maritime transport of passengers, in thousand passengers
eu_mtp_pas_emb	Maritime transport of passengers disembarked, in thousand passengers
eu_troad_cnl	Navigable canals, in kilometers
eu_troad_mway	Navigable canals, in kilometers
eu_troad_rd_oth	Motorways, in kilometers
eu_troad_riv	Other roads, in kilometers
eu_troad_rl	Navigable rivers, in kilometers
eu_troad_rl_elc	Total railway lines, in kilometers
eu_troad_rl_tge2	Electrified railway lines, in kilometers

eu_vs_bus_tot	Total number of motor coaches, buses and trolley buses
eu_vs_car	Total number of motor coaches, buses and trolley buses
eu_vs_lor	Total number of passenger cars
eu_vs_moto	Total number of lorries
eu_vs_spe	Total number of motorcycles
eu_vs_tot_x_tm	Total number of special vehicles
eu_vs_trc	Total number of all vehicles (except trailers and motorcycles)
eu_vs_trl_strl	Total number of road tractors
eu_vs_utl	Total number of trailers and semi-trailers
eu_epry2564f	Participation rate in education and training (last 4 weeks), females
eu_epry2564m	Participation rate in education and training (last 4 weeks), males
eu_epry2564t	Participation rate in education and training (last 4 weeks), total
Region_name	Nuts name
lifeexpectancy	Life expectancy
ghg	Emissions of total greenhouse gases
fossil	Emissions of fossil CO2