FORS PAPERS
Working paper series

Going beyond the single item: deriving and evaluating a composite subjective wellbeing measure in the Swiss Household Panel

Gondek, D.*[1], García Garzón, E.[2], Sánchez-Mira, N.[3], Vandecasteele, L.[4], Steinmetz, S.[4], Voorpostel, M.[1]

[1]FORS Swiss Centre of Expertise in the Social Sciences

[2]Camilo José Cela University, Faulted Salud

[3]University of Neuchâtel, Institute of Sociology

[4]University of Lausanne, Institute of Social Sciences

FORS Working Paper 01-2024

FORS Working Paper series

The FORS Working Paper series presents findings related to survey research, focusing on methodological aspects of survey research or substantive research. Manuscripts submitted are papers that represent work-in-progress. This series is intended to provide an early and relatively fast means of publication prior to further development of the work. A revised version might be requested from the author directly.

Further information on the FORS Working Paper Series can be found on www.forscenter.ch

How to cite this document:

Gondek, D., García Garzón, E., Sánchez-Mira, N., Vandecasteele, L., Steinmetz, S., Voorpostel, M. (2024). Going beyond the single item: deriving and evaluating a composite subjective wellbeing measure in the Swiss Household Panel. FORS Working Paper Series, paper 2024-1. Lausanne: FORS. DOI: 10.24440/FWP-2024-01

Acknowledgements

This paper is part of the research project "The Covid generation: Identifying risks and protective factors for young people's pathways through the Covid-19 pandemic in Switzerland" devoted to examining drivers of wellbeing during the pandemic and supported by Swiss National Science Foundation (request nr. 408040_210152). We thank the reviewers - Samuel Arthers, Mégane Bollenrücher, Oliver Lipps – for their insightful comments.

ISSN 1663-523x (online)

FORS, c/o University of Lausanne, Géopolis, 1015 Lausanne, Switzerland,
paperseries@fors.unil.ch

SUMMARY

The Swiss Household Panel (SHP) is an invaluable source of knowledge about wellbeing at the population level and its changes over time in Switzerland, allowing for cross-country comparisons. However, researchers using SHP data have been inconsistent in their choice and use of wellbeing indicators, making comparability of findings across studies difficult. With this guide, our aim was to derive an aggregate measure that maximises the SHP's potential to examine multiple dimensions of wellbeing and examine its validity and reliability. This will help researchers to make more informed decisions when using wellbeing measures in the SHP. This study was theoretically guided by the seminal work of Ed Diener on subjective wellbeing. Due to the availability of the measures over time, we focused on affect (emotional measures) and life satisfaction (cognitive measures).

We assessed the factorial structure and internal reliability of the wellbeing indicators available in the SHP and tested their measurement invariance across age groups, periods, gender, questionnaire languages, and survey modes. We demonstrated that combining single items in the SHP can derive a psychometrically robust wellbeing measure. Although an overall score of wellbeing combining all items into one indicator showed satisfactory internal reliability, such a one-dimensional measure should be used with caution, as our findings suggest that wellbeing as it is operationalized using the items available in the SHP is not a unidimensional construct. Instead, we recommend using two subscales that should be analyzed separately: 1) positive affect and life satisfaction, and 2) negative affect. However, caution is needed when age or language groups are compared, as certain items behaved differently across groups.

This guide provides a step-by-step approach to developing a wellbeing measure that combines single items from different batteries and rigorously assesses its statistical properties. In this way, it can inform researchers using the SHP data on how to move beyond using separate items to construct a wellbeing measure consisting of two dimensions. Furthermore, bringing more consistency to analyzing wellbeing using the SHP will facilitate comparability and help interpret effect sizes.

Keywords: wellbeing, measure, positive affect, life satisfaction, negative affect

1. Introduction

In Switzerland, the Swiss Household Panel (SHP) is the only large longitudinal data source that allows monitoring within people's wellbeing over time since 1999 (Tillmann, Voorpostel, Antal, Dasoki, Klaas, Kuhn, Lebert, Monsch & Ryser, 2021). The dataset includes individual items capturing components of subjective wellbeing, including affect (emotional measures) and life satisfaction (cognitive measures). Subjective wellbeing is defined as "a person's cognitive and affective evaluations of his or her life" (Diener et al., 2002, p. 63). According to the author's tripartite model (Diener, 1984), which serves as a theoretical framework for our study, wellbeing has three components: life satisfaction, positive affect, and negative affect. The affective components of wellbeing have been studied less often than life satisfaction, despite being of equal importance (REF). One potential reason for this is that emotional responses have been shown to be short lived and fluctuating, whereas life satisfaction is cognitively appraised, hence not as susceptible to short-term influences of live events (Gilman, Huebner, & Laughlin, 2000; Luhmann, 2017).

The questionnaire items capturing wellbeing in SHP do not come from a validated scale measuring wellbeing, such as for instance, the Positive and Negative Affect Schedule (PANAS-SF) (Watson, Clark & Tellegen, 1988). Individual items capturing aspects of wellbeing have been designed or adapted from various sources for the purpose of the survey. The psychometric properties of these items, in any aggregated form, or their interrelations have not been previously explored. This provides motivation for our study, which is to develop a composite measure of subjective wellbeing. This could help researchers make an informed choice of wellbeing measure and improve comparability of research using the SHP.

Wellbeing has been widely studied using the SHP. According to the FORS publications directory (https://forscenter.ch/publications/), a total of 47 studies using the SHP and including the keywords "wellbeing" (or "well-being"), "mental health" and "life satisfaction" in the title were published between 2018 and 2022. These studies have examined the effects of a variety of factors on wellbeing, such as the regularisation of migrant workers (Burton-Jeangros et al., 2021), chronic physical health (Debnar et al., 2021), or reduced employment (Schröder, 2020). These studies have selected and combined these items in different ways deriving aggregated measures of positive and negative affect (Dawson-Townsend, 2019), or using individual items (Barbuscia & Comolli, 2021; Chesters et al., 2021; Comolli et al., 2021; Henning et al., 2023). The same items of the questionnaire have been referred to as measuring mental health (Barbuscia & Comolli, 2021), depressive symptoms (Barbuscia & Comolli, 2021), wellbeing (Chesters et al., 2021), low mood (Węziak-Białowolska, 2016), as well as life satisfaction (Henning et al., 2023). Some of these differences may result from researchers coming from different disciplines, such as psychology, sociology or economics. However, the key issue is that criteria for selecting items are rarely explicit. Statistical properties of aggregated measures are rarely provided. Using all relevant items is often impractical due to multiple and potentially divergent results across outcomes, which makes interpreting findings difficult.

Another complication of research using subjective wellbeing in general is that it varies greatly how researchers conceptualise the factorial structure of subjective wellbeing, treating wellbeing as an overall composite, keeping the three components as separate (Lucas et al. 1996), or as configurations of components (Busseri & Sadava, 2011). Numerous studies suggest wellbeing is unlikely unidimensional. Therefore, combining its several components into one composite indicator may be deemed poor practice (Daniel-González et al., 2020; Jovanović, 2015; Metler & Busseri, 2017). More recently, an increasing number of studies have shown a hierarchical structure, with wellbeing as a higher order (or a superordinate) factor comprising of subcomponents, referred to as lower order (or subordinate) factors (Daniel-González et al., 2020; Jovanović, 2015; Metler & Busseri, 2017). As evidence on the factorial structure of wellbeing is largely mixed, it has been recommended that researchers specify and justify their underlying assumptions regarding their structural model for a measure of subjective wellbeing in each study (Busseri & Sadava, 2011). However, it is also important that information on the statistical properties of the used measure is also provided due to potential sample differences. As emphasized by Diener (1984), it is an empirical, rather than theoretical, question of how the components of wellbeing relate to each other, and of key interest is their relationship with other variables.

The aim of this paper is to develop a composite measure of subjective wellbeing to improve the efficiency and comparability of research using the SHP. We assess and examine the statistical properties, as well as offer guidelines on how to construct and use this composite measure in future studies. This should save valuable time and make researchers' decisions about wellbeing measures more informed. Furthermore, using a consistent measure of wellbeing can make research using the SHP more comparable, which will facilitate interpreting for instance effect sizes across studies.

2. Methodology

2.1 Data

The SHP is a household-based panel study that collects annual information on various aspects of life from each person living in the household at the time of the interview (Tillmann et al., 2021; SHP Group, 2023). Its key strength is that it is representative of private households in Switzerland. Since 1999, the SHP has been a major source of information on (changes in) health, family, work, attitudes, political participation, and migration.

After the initial sample that started in 1999, refreshment samples were added in 2004, 2013 and 2020. For most of the analyses, we used data from wave 2020, that is, including the latest refreshment sample. The questionnaire was administered mainly by Computer Assisted Telephone Interviewing (CATI) (71.2%), with Computer Assisted Web Interviewing (CAWI) (28.8%). Eligible participants included those aged 14 years or older who participated in the SHP in 2020, as from this age participants completed the individual questionnaires (n = 15882). Due to missing information on selected wellbeing indicators, the analytical sample was reduced to 15233 individual participants across the analyses.

For one of the analyses, which was to test the consistency of the constructs across time (4.6 Measurement invariance across various groups), it was necessary to use data collected at multiple time points. To ensure that we do not conflate within and between person effects, the same participants could only be observed once. That is, data from 2006 included SHP original sample and refreshment sample one, 2013 included refreshment sample two and 2020 included refreshment sample three (n total = 13702).

2.2 Measures

Following the framework proposed by Diener et al. (2002), we selected 11 potentially relevant items that measured various aspects of life satisfaction (5 items) as well as positive (2 items) and negative affect (4 items). The key inclusion criterion was for the items to be part of the core annual questionnaire for all respondents. This allows deriving a measure of wellbeing that can be used across most of the SHP waves and for all respondents of the individual questionnaire on an annual basis. Therefore, we did not include modular questions that are asked less frequently or to only part of the sample, for instance those related to satisfaction with employment or romantic relationships or the entire Diener's life satisfaction scale. The selected items are described in more detail in Table 1. The response options for all items range from 0 to 10, with a higher score indicating greater satisfaction in each life domain or more frequent positive and negative affect. We treated the items as continuous, rather than ordinal as, due to a relatively wide range of response options (11 points), where only extreme ends of the scale were assigned a qualitative label (REF).

Table 1. Selected individual items measuring different aspects of wellbeing in the SHP.

Questions	Dimension	Variable code	Response options	Years collected
Life satisfaction: In general, how satisfied are you with your life?	Satisfaction with various life domains	C$$PC44	0 (not at all satisfied) – 10 (completely satisfied)	Since 2000
Health satisfaction: How satisfied are you with your state of health"?	Satisfaction with various life domains	C$$PC02	0 (not at all satisfied) – 10 (completely satisfied)	Since 1999
Financial satisfaction: Overall how satisfied are you with your financial situation?	Satisfaction with various life domains	P$$I01	0 (not at all satisfied) – 10 (completely satisfied)	Since 1999
Relationships satisfaction: How satisfied are you with your personal, social and family relationships?	Satisfaction with various life domains	P$$QL04	0 (not at all satisfied) – 10 (completely satisfied)	Since 2001
Leisure time satisfaction: How satisfied are you with your leisure time activities?	Satisfaction with various life domains	P$$A06	0 (not at all satisfied) – 10 (completely satisfied)	Since 1999
Energy and optimism: Are you often plenty of strength, energy and optimism?	Positive affect	C$$PC18	0 (never) – 10 (always)	Since 2000
Joy: How frequently do you generally experience the following emotions?	Positive affect	P$$C47	0 (never) – 10 (always)	Since 2006
Anger: How frequently do you generally experience the following emotions?	Negative affect	P$$C48	0 (never) – 10 (always)	Since 2006
Sadness: How frequently do you generally experience the following emotions?	Negative affect	P$$C49		
Worry: How frequently do you generally experience the following emotions?	Negative affect	P$$C50		
Anxiety and depression: Do you often have negative feelings such as having the blues, being desperate, suffering from anxiety or depression?	Negative affect	C$$PC17	0 (never) – 10 (always)	Since 1999

3. Analytical strategy

Table 2, included at the end of the section, includes a summary of the analytical steps. The code for all analyses can be found at https://osf.io/vnzcw/?view_only=ffad7d69ae24416692bbdb413363f185.

3.1 Descriptive information and exploratory graph analysis

First, we calculated descriptive statistics for the 11 items, such as mean, standard deviation, correlation, and distribution. Second, we described the relationships between the wellbeing indicators in an exploratory manner and examined whether they fell on any common clusters. To do this, we used exploratory network analysis (EGA), which is based on a network approach (Costantini & Epskamp, 2017). We opted for EGA, instead of an exploratory factor analysis due to its improved accuracy in estimating the correct number of factors and the better visualization of complex interrelationships between variables (Golino & Christensen, 2022; Christensen & Golino, 2021). We also used EGA to potentially reduce the number of relevant items if they did not fall on the same factor. In addition, this analytical step informed the subsequent analysis, in which we further tested the factorial structure of the measure (see 3.2 Confirmatory factor analysis).

Network analysis is typically presented graphically, consisting of nodes (circles representing variables) and edges (lines signifying relationships between nodes, here variables) (Epskamp & Fried, 2018). EGA not only reveals how items interact with each other, but also quantifies the importance of items relative to others (centrality measures), which are roughly equivalent to factor loadings (Christensen & Golino, 2021). The correlations between items are partial (controlled for all other items) and regularised (shrunk towards null) (Epskamp & Fried, 2018). The shrinkage was done using the EBIC-based graphical lasso regularization parameter, to encourage simpler models, with fewer parameters (Epskamp & Fried, 2018). The possible dimensions (or clusters) of items were identified using the Walktrap algorithm, with items assigned into clusters with small intra and larger inter-community distances (Pons & Latapy, 2006). We evaluated stability of the clusters by re-estimating their number with 1000 bootstraps using a parametric approach (Christensen & Golino, 2021). The analyses were conducted in R (v.4.3.1) (Team, 2021), using the EGA and bootEGA functions, available in the EGAnet package (v.1.2.3) (Golino & Christensen, 2022) and the network was displayed with the qgraph package (v.1.9.5) (Epskamp et al., 2012).

3.2 Confirmatory factor analysis

Third, we tested the factorial structure of the items in a confirmatory fashion using confirmatory factor analysis (CFA). While EGA was an exploratory, data-driven, exercise where we did not assume any specific data-generating process, the models for the CFA were theoretically driven, after accounting for the information gained through the EGA. That is, we compared models that assume unidimensional structure, hierarchical or two-subcomponent structure of subjective wellbeing. We did not assess the fit of model with three subcomponents, as exploratory analysis showed that positive affect and life satisfaction fell on one factor. Besides, positive affect was measured only by two items, hence there may be little value in creating a separate factor capturing common variance of only these two items.

This provided information on whether the items can be aggregated as a total wellbeing score, or across its subdomains.

We compared four models, which are typically examined in the context of wellbeing measurement: 1) unidimensional, 2) correlated factors, 3) bifactor, 4) bifactor S-1. The unidimensional model assumes that the items capture a single construct and hypothesizes that a single factor is sufficient to explain the variance in all indicators of wellbeing, hence an aggregated (by summing up or averaging) overall wellbeing score could be used (Gustafsson & Åberg-Bengtsson, 2010). The correlated factors model hypothesizes that the items capture separate, but related components (Brown, 2015). A strong fit of such a model would suggest that scores aggregated at the subcomponent level of wellbeing could be preferred (e.g., capturing negative affect). In the bifactor model, each item loads onto a general factor, and factors representing subcomponents, where each subcomponent accounts for unique variance beyond the variance capturing overall wellbeing (Holzinger & Swineford, 1937). Hence, the bifactor model allows to capture a "purer" representation of the subcomponent of wellbeing, after partitioning out variance capturing other components of wellbeing (Eid et al., 2017). A strong fit of this model would suggest that there is a broad general wellbeing factor that underlies all wellbeing indicators as well as conceptually narrower specific factors, such as negative or positive affect. These specific factors could be of theoretical interest to researchers and would need to be modelled as latent variables, instead of being represented by aggregated scores across items.

The bifactor model has been widely used in psychopathology and wellbeing research, but it has also been heavily criticized due to the difficulty of interpreting both the general factor and factors corresponding to subcomponents, as well as convergence problems and anomalous results (Heinrich et al., 2023). Therefore, we also fit the bifactor S-1 model. The difference between this model and a more traditional bifactor model is that in the bifactor S-1 model the researcher explicitly choses a subset of items (reference domain) that load exclusively on the general factor. This is in opposition to the symmetrical bifactor model, in which all items determine the composition of the general factor. In the bifactor S-1 model, the meaning of the general factor is *defined a priori*. For instance, if the researcher decides that the negative affect serves as a general factor, other subcomponents will represent the unique variance that is not shared with negative affect. In this situation, it is easier to interpret the meaning of these subcomponents (e.g., positive affect and life satisfaction without the shared variance with negative affect).

We interpreted the fit of each model according to criteria proposed by Schermelleh-Engel et al., (2003), where fit was deemed as acceptable if: the Root Mean Square Error of Approximation (RMSEA) ≤ 0.08 the Standardised Root Mean Residual (SRMR) ≤ 0.10 (Schermelleh-Engel et al., 2003), the Goodness of Fit Index (CFI) ≥ 0.95 and the Tucker–Lewis Index (TLI) > 90 (Bentler & Bonett, 1980). However, it is commonly emphasised that fit indices should not solely determine whether a given model is considered in useful in practice. Simulation studies showed that deterministic choice of cut-off values may lead to either rejection of strong models or acceptance of poor ones, depending on various aspects of the data, such as sample size (Beauducel & Wittmann, 2005; Little et al., 2002).

The CFA models were conducted using the lavaan package in R (Rosseel, 2012).

3.3 Internal consistency

Fourth, we evaluated the internal consistency across the items, to assess to what extent the items measure the same construct. For internal consistency, the negative affect items were reversed so the higher scores reflect higher wellbeing. We reported the omega coefficient in addition to the typically used alpha coefficient. The omega coefficient is an alternative measure of reliability, robust in cases where loadings are not equal for all items and the construct is not unidimensional, which is often the case with wellbeing (Zhang & Yuan, 2016). We estimated two omega coefficients proposed by McDonald that account for the hierarchical structure of the measure – omega hierarchical (ω^h) and total (ω^t) (Zhang & Yuan, 2016). ω^h allows to evaluate the importance and reliability of the general factor of a measure, while ω^t is an estimate of the total reliable variance in a measure. Reliability was assessed to provide additional insight into total scoring. However, if unidimensionality was clearly not supported, alpha coefficient results should be treated with caution. Omega coefficients allow for multidimensional hierarchical structure of a test. The coefficients were obtained by using the *reliability* function from the *semTools* R package (v.0.5-6), after running the factor models (Jorgensen, Pornprasertmanit, Schoemann & Rosseel 2022).

3.4 Correlation between aggregated (sum) and factor scores

Fifth, we calculated correlation coefficients between sum and factor scores of overall wellbeing and its subdomains. This provides us with information on how similar the scores on wellbeing and its subcomponents are when aggregated or derived from factor scores. If the correlations are very high, similar results would be expected in subsequent analysis where wellbeing is used as an exposure or outcome.

The sum scores were derived by aggregating the items, where factor scores were obtained from each of the four CFA models using the Bartlett method (Bartlett, 1937). The factor scores were standardised, with a mean of 0 and variance of 1. Factor scores, through weights corresponding to factor loadings, reflect the potentially differential contribution of each item to the total score (McNeish & Wolf, 2020). In case of simple sum (or average) scores, the weighting of each item is equal (McNeish & Wolf, 2020). That is, each item contributes equal information (or is equally important for) to the overall concept. It is a strong and rather unrealistic assumption. Factor scores have been argued to be a more accurate representation of the rank order of each participant, however, they have their disadvantages (McNeish & Wolf, 2020; Widaman & Revelle, 2023). For instance, they make the statistical analysis more complex, where ideally the measurement model of the latent variable (i.e., wellbeing) and the structural model of interest (e.g., examining an association between gender and wellbeing) are included in the same overall model simultaneously. This may result in convergence problems. Sum scores are often preferred, as they do not require latent variable modelling. In addition, sum scores provide a greater comparability of results, as they do not depend on the sample from which they are derived (i.e., each participant would have the same score regardless of other individuals in the study). If sum and factor scores correlate very strongly, they are likely to lead to highly comparable results, in which case, a simpler approach using sum scores could be chosen (Widaman & Revelle, 2023).

3.5 Association between sum/factor scores and covariates

Sixth, we described the associations of the sum or factor scores with a range of commonly used socio-demographic covariates including age (in years), gender (men, women), language

(French, German, Italian), and mode of data collection (CATI, CAWI). The coefficients were expressed as unstandardised (B) for sum scores and standardised (β) for factor scores. The aim of the analysis is to compare the results across both types of scores, to see whether the choice of the score may lead to different results, when as in this case, wellbeing is used as an outcome. For simplicity, we first derived the factor scores and then included them as outcomes in the regression analysis, instead of using measurement and structural models simultaneously. All covariates were simultaneously included in the models, hence mutually adjusting for each other.

3.6 Measurement invariance across various groups

Seventh, we tested measurement invariance across age groups (14-35, 36-65, >65), period (2006, 2013, 2020), gender (men, women)[1], language (French, German, Italian), mode of collection (CATI, CAWI) using multigroup CFA. This was to assess whether the interpretation of the items is equivalent across groups. We examined three forms of invariance: (1) configural, equivalence of model form; (2) metric (weak factorial), equivalence of factor loadings; (3) scalar (strong factorial), equivalence of factor loadings and intercepts. The measurement invariance is typically tested by comparing fit of increasingly stringent models, starting with a configural model. That is, when the fit of the metric model is at least as good as the configural one, weak invariance is achieved, while when the fit of the scalar model is at least as good as the metric one, the invariance is assumed to be strong.

There is no consensus about the best fit indices or cut-off values for assessing whether the models sufficiently differ (or not) in their fit, to judge the level of invariance. Among the most used criteria are a -.01 change in CFI, paired with changes in RMSEA of .015 and SRMR of .030 (for metric invariance) or .015 (for scalar invariance) (Chen, 2007). We interpret our findings following these suggested cut-offs as well as the overall fit of the models.

[1] SHP does not include information on non-binary gender identification.

Table 2. A summary of the analytical steps.

Analytical step	Aim
1. Descriptive information	To provide descriptive statistics for the 11 items, such as mean, standard deviation, correlation, and distribution.
2. Exploratory graph analysis	To describe the relationships between the wellbeing indicators in an exploratory manner and examined whether they fell on any common clusters.
3. Confirmatory factor analysis	To test the factorial structure of the items in a confirmatory fashion.
4. Internal consistency	To assess to what extent the items measure the same construct.
5. Correlation between aggregated (sum) and factor scores	To provide information on how similar the scores on wellbeing and its subcomponents are when aggregated or derived from factor scores.
6. Association between sum/factor scores and covariates	To describe the associations of the sum or factor scores with a range of commonly used socio-demographic covariates including age, gender, language, and mode of data collection.
7. Measurement invariance across various groups	To assess whether the interpretation of the items is equivalent across groups, including age, period, gender, language, and mode of data collection.

4. Results

4.1 Descriptive information and exploratory graph analysis

The mean scores across the study sample on various domains of life satisfaction ranged from 7.31 (satisfaction with financial situation) to 8.23 (satisfaction with personal relationships) (see Table 3). The mean values for positive affect items varied between 7.06 (energy and optimism) to 7.48 (joy). Among the items of negative affect, the mean scores ranged from 2.18 (depression, blues, and anxiety) to 3.88 (anger). There was a much greater variability, as indicated by standard deviation, in scores on negative affect items in relation to the population mean than for positive affect and life satisfaction (see Table 3).

Table 3. Descriptive information about the wellbeing items.

Item	**Mean (SD)**
Satisfaction with life in general	8.14 (1.40)
Satisfaction with health status	7.84 (1.73)
Satisfaction with financial situation	7.31 (2.16)
Satisfaction with personal relationships	8.23 (1.48)
Satisfaction with leisure activities	7.78 (1.98)
Energy and optimism	7.06 (1.86)
Joy	7.48 (1.35)
Anger	3.88 (1.97)
Sadness	3.44 (2.02)
Worry	3.21 (2.33)
Depression, blues, and anxiety	2.18 (2.18)

N = 15,324; SD = standard deviation; The values ranged from 0 to 10 on all items.

Source: SHP (2020) author's own calculations.

The positive affect and satisfaction measures were somewhat left skewed with few individuals with low scores, whereas negative affect items were slightly right skewed with a low proportion of participants with high scores (see Figure S1 in the Appendix).

Overall, the correlations between the items were weak to moderate (see Figure 1), ranging from -0.12 (between joy and anger) to 0.58 (between sadness and worry). The positive affect and life satisfaction items correlated positively with each other and negatively with negative affect. Depression, blues and anxiety moderately correlated with multiple other items ranging from -0.45 (life satisfaction) and 0.54 (sadness). Financial satisfaction appeared to have the weakest correlations with other items, whereas life satisfaction had the strongest and most consistent correlation with other items (ranging from -0.26 with worry to 0.50 with joy).

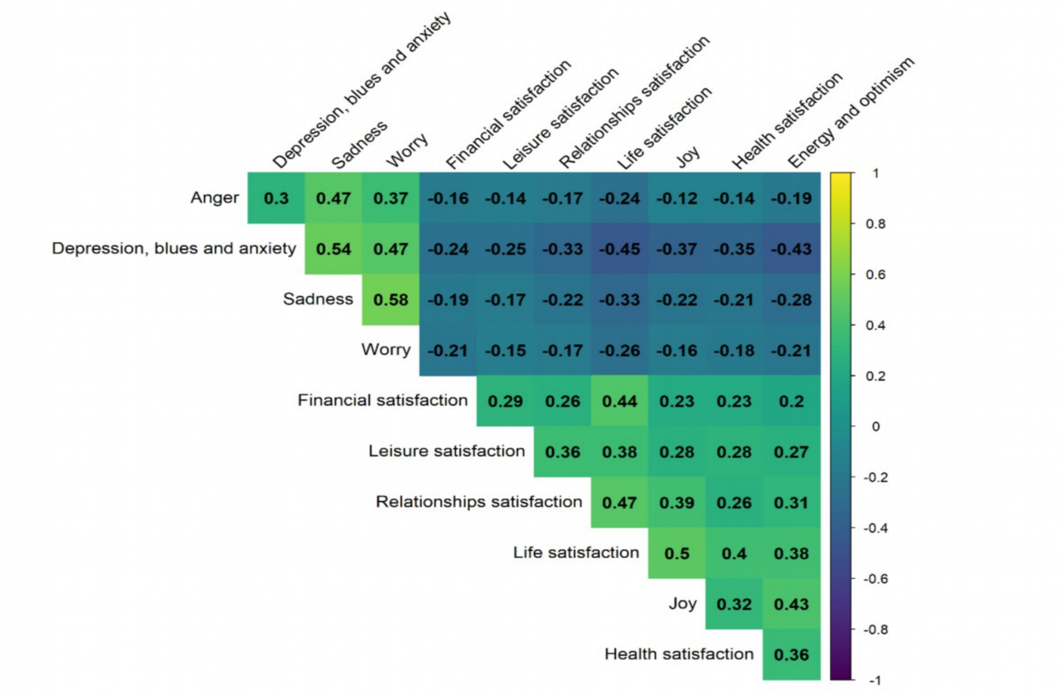
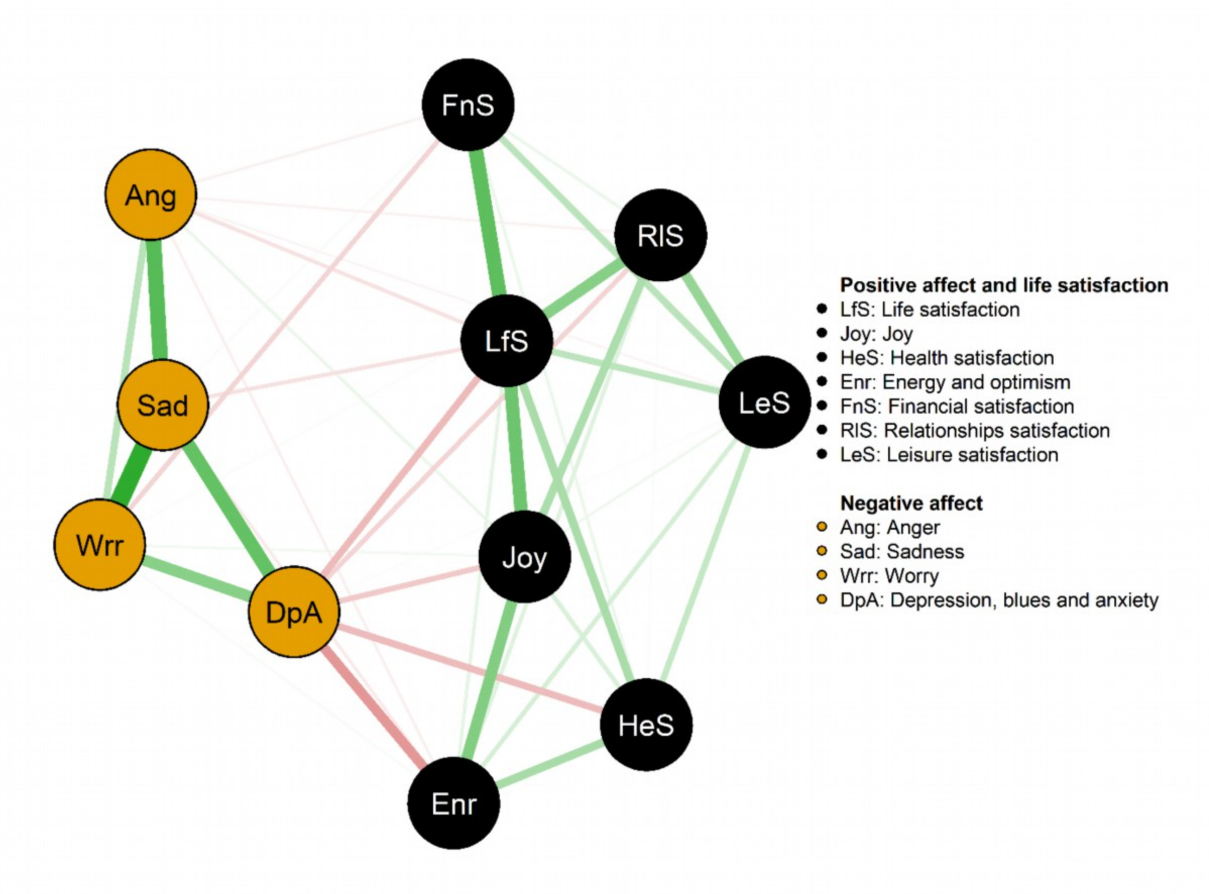


Figure 1. Correlation matrix of indicators of wellbeing.

Source: SHP (2020) author's own calculations.

Using the exploratory graph analysis, we arrived at two clusters: one grouping items of positive affect and life satisfaction, and another of negative affect (see Figure 2). As shown by the green, relatively thick lines, life satisfaction was positively and strongly connected with most items within the cluster. This was confirmed by the high centrality measure (0.47). Sadness was central for the negative affect cluster (strength centrality = 0.50). This implies life satisfaction constitutes the core element of positive affect and life satisfaction cluster, and sadness of negative affect cluster. Namely, these items are most important for those two clusters, bridging other items together.

As satisfaction with the financial situation of the household did not appear to be well-connected with other items of positive affect and life satisfaction (strength centrality = 0.19), we dropped this item from all following analyses. The depression and anxiety item was strongly negatively related to multiple items from the dimension of positive affect and life satisfaction, reflected by strength centrality within the dimension of positive affect and life satisfaction = -0.26.



procedure showing two dimensions in wellbeing indicators.

Source: SHP (2020) author's own calculations.

4.2 Confirmatory factor analysis

The visual representation of all the fitted factors is shown in Figure 3 (the factor loadings can also be found in Table S1 in the Appendix). The unidimensional model had an unsatisfactory fit according to most indices (e.g., CFI = 0.774, TLI = 0.709, RMSEA = 0.134, see Table 4), indicating that the wellbeing items cannot be reduced to one single dimension. This was expected considering the results of the EGA, which indicated presence of two wellbeing dimensions: 1) positive affect and life satisfaction, 2) negative affect. Therefore, subsequently, we fit a 2-factor model, where the factors were allowed to correlate. We expected the correlation, as the EGA indicated that some items across both dimensions are correlated (e.g., life satisfaction – depression and anxiety). The fit of the 2-factor model was substantially better than that of the unidimensional model, with the cut-offs typically considered on the verge of being acceptable (e.g., CFI = 0.920, TLI = 0.894, RMSEA = 0.081).

Next, both bifactor and bifactor S-1 had an excellent fit (see Table 4). In the bifactor model, all loadings apart from energy and optimism were marginal onto the dimension of positive affect and life satisfaction (0.05 – 0.33, see Figure 3 & Table S1 in the Appendix). The loading of energy and optimism was reversed to what was expected from the theory and found in the 1- and 2-factor models, reaching a relatively high value of -1.29. The results indicate that the residual positive affect was largely driven by the energy item and that the remaining positive items were sufficiently explained by the general well-being factor, without needing and additional factor. This phenomenon is known as factor collapse (Mansolf & Reise, 2016).

As the next step, we explicitly modelled the general factor as positive affect and life satisfaction (acting as a reference factor) in the bifactor S-1 model. In the bifactor S-1 model, the loadings of the negative affect were nearly identical to those in the bifactor model. This confirms the factor collapse in the bifactor model. In the bifactor S-1 model, negative affect represents the part of the domain that cannot be explained by the positive affect and life satisfaction, which serves as a reference general factor.

Table 4. Fit indices across the four factor models.

	Unidimensional	Two-factor (correlated)	Bifactor	Bifactor S-1
Chi square (# of parameters)	9689.739 (20)	3443.385 (21)	843.173 (30)	1552.117 (24)
CFI	0.774	0.920	0.981	0.964
TLI	0.709	0.894	0.965	0.948
RMSEA	0.134	0.081	0.046	0.057
SRMR	0.084	0.055	0.021	0.029

CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared error.

Source: SHP (2020) author’s own calculations.

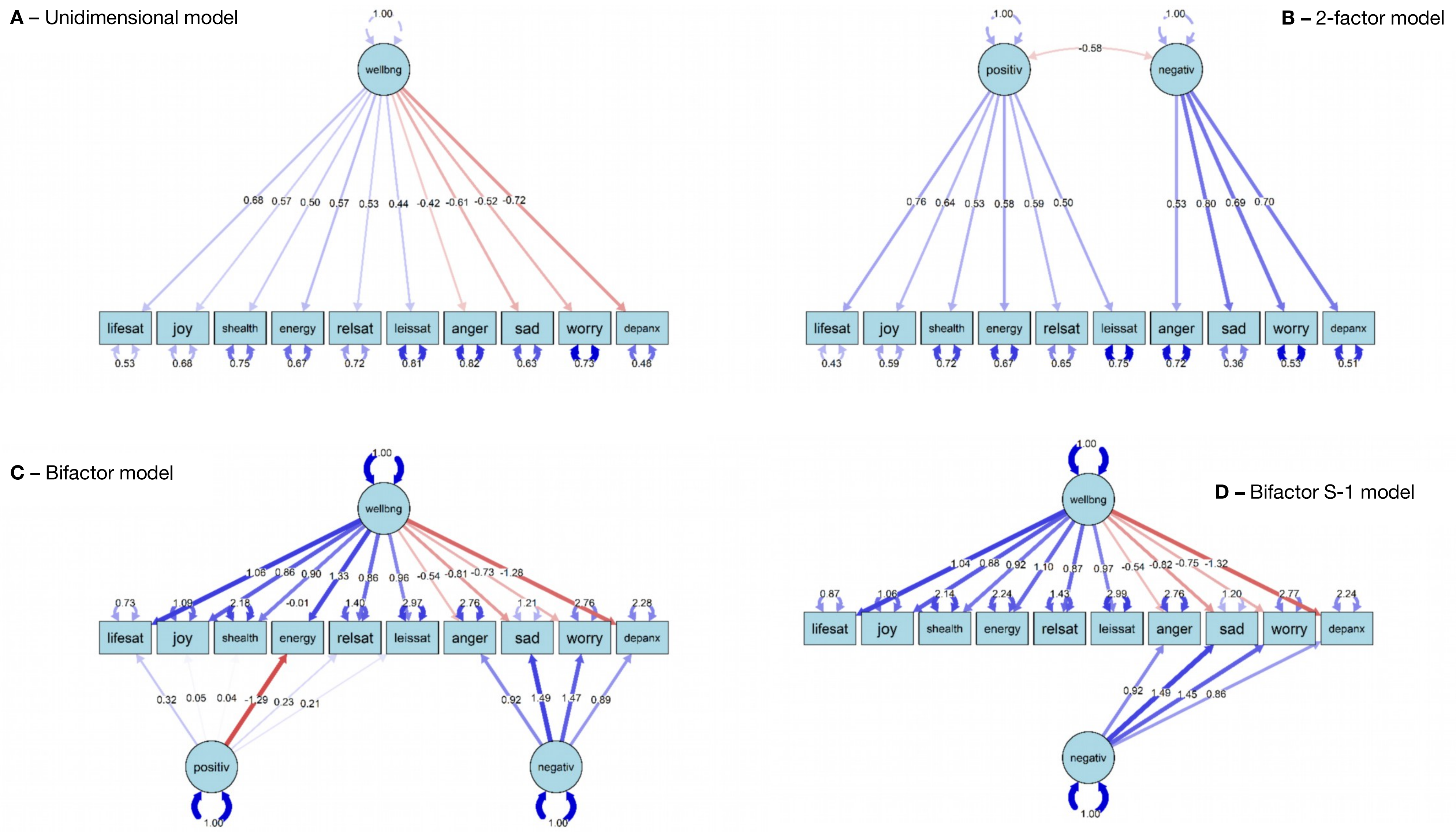


Figure 3. Graphical representation of the four factor models. The factor loadings can also be found in Table S1.

Source: SHP (2020) author's own calculations.

4.3 Internal consistency

The internal reliability across all wellbeing items was overall satisfactory. The alpha coefficient representing the unidimensional model was high (α = 0.81). The omega hierarchical (ω_h), based on the bifactor S-1 model, was 0.67. Thus, the proportion of variance in wellbeing that was due to a general wellbeing factor over and above the influence of effects that were specific to the negative affect was 0.67. The lower value of the omega hierarchical coefficient than alpha coefficient was potentially less biased due to the multidimensional nature of our scale.

4.4 Correlation between aggregated (sum) and factor scores

In this section we describe correlation between aggregated sum scores and factor scores derived from each factor model. This can inform about the extent to which results of an analysis could differ when using wellbeing on an aggregate level or its subcomponents, as an outcome, exposure or covariate.

The wellbeing sum score and factor score derived from the unidimensional model correlated nearly perfectly (Pearson's r = 0.99), and very highly with the general wellbeing factor from the bifactor and bifactor S-1 models (r = 0.89 and 0.91, respectively; see Figure 4). The correlation was also strong with both positive (r = 0.86) and negative affect (r = -0.85) sum scores. The wellbeing factor score derived from the bifactor model was almost perfectly correlated with the wellbeing factor score obtained from the bifactor S-1 model (r = 0.98), which provides support for collapsing the positive affect and life satisfaction factor in the bifactor model.

The sum score of positive affect and life satisfaction and the factor score derived from the 2-factor model also had a nearly perfect correlation (r = 0.97). This means that when these aggregated or factor scores are used in an analysis, as exposures, covariates or outcomes, the findings should be highly comparable. Hence, some researchers may opt for a simpler model using aggregated scores. The correlation between the sum score and factor score from the bifactor model was nearly zero (r = -0.07), which again suggests that the variance in the bifactor model shifted from the positive affect and life satisfaction factor to the general wellbeing factor. The correlation between the positive affect and life satisfaction and negative affect sum scores was negative and moderate (r = -0.45). Hence, one could expect to obtain different results when using these variables in their analyses.

Finally, the negative affect sum score was very highly correlated with the factor score from the 2-factor model (r = 0.98). Again, this would suggest a little advantage of using factor score over a simpler sum score. The correlation between the negative affect sum score and factor score derived from the bifactor and bifactor S-1 models was 0.85. The factor scores of the negative affect from the bifactor and bifactor S-1 models represent negative affect without the variance shared with the positive affect and life satisfaction. Therefore, it could be expected that the results obtained using such a conceptualisation of negative affect could differ from those when more conventional sum or factor scores capturing negative affect are used.

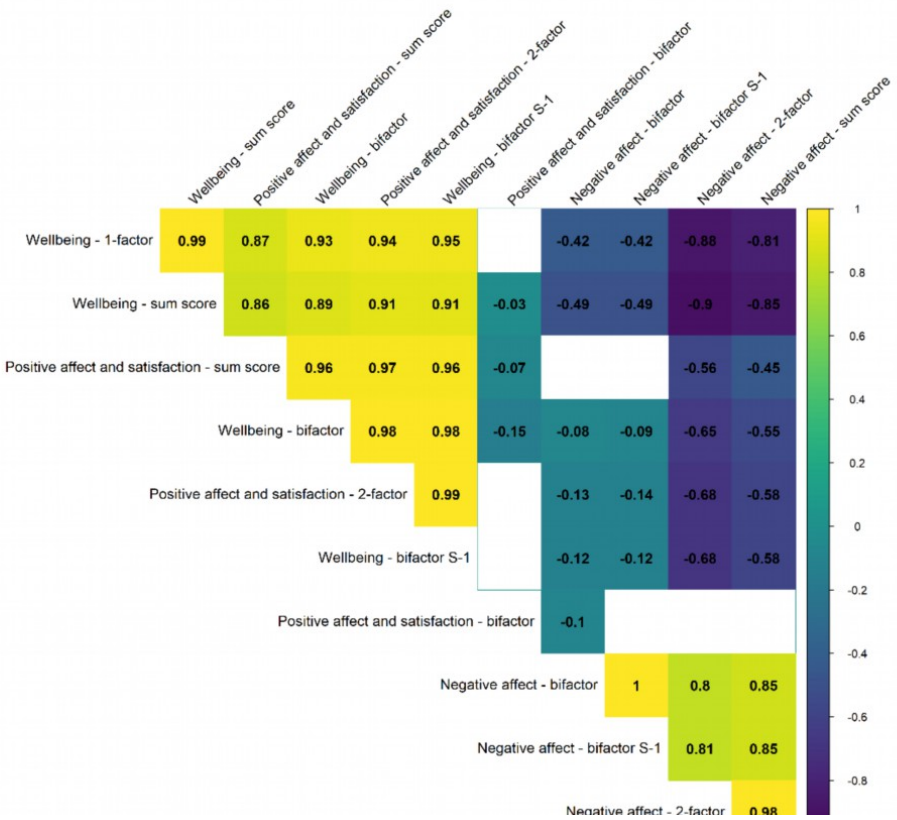


Figure 4. Correlation between sum and factor scores obtained from the four factor models.

Note. Blank spaces indicate lack of correlation.

Source: SHP (2020) author's own calculations.

4.5 Association of sum/factor scores with covariates

The previous section gave us an idea about how likely differing results can be obtained when either sum or factor scores are used. In this section, we use these scores as outcomes and regress them on various basic sociodemographic exposures. This would further show whether substantial differences in results are likely when one decides to use sum or factor scores in their substantial analysis.

The associations with covariates were highly comparable for wellbeing derived as either the sum score or factor score from the unidimensional model (see Table 5), as expected from the high correlations between the wellbeing scores. There were some differences in the association between wellbeing scores and covariates including gender and questionnaire language. The differences between groups being smaller when factor scores from bifactor models were used. These differences were likely due to the factor scores reflecting the subdomain of wellbeing – the positive affect and life satisfaction (by design in the bifactor S-1 model).

The associations between positive affect and life satisfaction and the covariates are highly comparable across scores derived from different models. The minor differences observed were a somewhat greater standardised difference in scores by mode of data collection derived from a bifactor model than a sum score (-0.03, -0.06 to -0.00 vs -0.13, -0.16 to -0.11)

For negative affect, the associations with covariates were again largely comparable for scores derived from different models. Some differences were found for the associations with age. The scores from the bifactor models were weakly and positively linked with age (e.g., 0.01, 0.00 to 0.03 in bifactor S-1 model), while the sum score and factor score from 2-factor model had a modest negative association with age (e.g., -0.03, -0.04 to -0.01 for sum score). This supports the notion that the negative affect factor from the bifactor models may capture a largely different construct.

Table 5. The association between wellbeing sum and factor scores and covariates.

	Wellbeing (unidimensional model)			Wellbeing (bifactor model)			Wellbeing (bifactor S-1 model)			Wellbeing (sum score)		
	B	CI95% low	CI95% high	B	CI95% low	CI95% high	B	CI95% low	CI95% high	B	CI95% low	CI95% high
Gender (reference: men)												
Women	-0.12	-0.14	-0.10	-0.03	-0.05	-0.01	-0.04	-0.06	-0.01	-0.13	-0.15	-0.11
Age (in years)	0.03	0.01	0.04	0.04	0.03	0.06	0.03	0.02	0.05	0.03	0.01	0.05
Language of interview (reference: French)												
German	-0.02	-0.07	0.03	-0.04	-0.09	0.01	-0.04	-0.10	0.01	0.00	-0.05	0.05
Italian	-0.25	-0.28	-0.22	-0.10	-0.14	-0.07	-0.13	-0.17	-0.10	-0.27	-0.30	-0.23
Mode of interview (reference: CATI)												
CAWI	-0.12	-0.14	-0.09	-0.10	-0.13	-0.07	-0.11	-0.13	-0.08	-0.14	-0.16	-0.11
	Positive affect and life satisfaction (2-factor model)			Positive affect and life satisfaction (bifactor model)			Positive affect and life satisfaction (bifactor S-1 model)			Positive affect and life satisfaction (sum score)		
	B	CI95% low	CI95% high	B	CI95% low	CI95% high	β	CI95% low	CI95% high	β	CI95% low	CI95% high
Gender (reference: men)												
Women	0.01	-0.01	0.03	0.04	0.02	0.06	NA	NA	NA	0.00	-0.02	0.02
Age (in years)	0.03	0.01	0.05	-0.01	-0.02	0.01	NA	NA	NA	0.03	0.01	0.04
Language of interview (reference: French)												
German	-0.03	-0.08	0.03	0.03	-0.02	0.08	NA	NA	NA	-0.05	-0.10	0.01
Italian	-0.09	-0.12	-0.05	-0.12	-0.15	-0.08	NA	NA	NA	-0.07	-0.11	-0.04
Mode of interview (reference: CATI)												
CAWI	-0.11	-0.13	-0.08	-0.03	-0.06	0.00	NA	NA	NA	-0.13	-0.16	-0.11
	Negative affect (2-factor model)			Negative affect (bifactor model)			Negative affect (bifactor S-1 model)			Negative affect (sum score)		
	B	CI95% low	CI95% high	B	CI95% low	CI95% high	β	CI95% low	CI95% high	β	CI95% low	CI95% high
Gender (reference: men)												
Women	0.24	0.22	0.26	0.26	0.24	0.29	0.27	0.25	0.29	0.23	0.20	0.25
Age (in years)	-0.01	-0.03	0.00	0.02	0.00	0.03	0.01	0.00	0.03	-0.03	-0.04	-0.01
Language of interview (reference: French)												
German	-0.01	-0.06	0.04	-0.04	-0.09	0.01	-0.04	-0.09	0.01	-0.05	-0.10	0.00
Italian	0.40	0.37	0.43	0.42	0.39	0.45	0.41	0.37	0.44	0.38	0.35	0.42
Mode of interview (reference: CATI)												
CAWI	0.10	0.08	0.13	0.07	0.04	0.09	0.07	0.04	0.09	0.10	0.08	0.13

CATI = Computer Assisted Telephone Interview; CAWI = Computer Assisted Web Interview; NA = not applicable (positive affect and life satisfaction was used as a reference category).
Note. The coefficients were expressed as unstandardised (B) for sum scores and standardised (β) for factor scores.

Source: SHP (2020) author's own calculations.

4.6 Measurement invariance across various groups

The final step of our analysis was to test measurement invariance across age (14-35, 36-65, >65), period (2006, 2013, 2020), gender (men, women), language (French, German, Italian), and mode of data collection (CATI, CAWI). We tested MI of three models: 2-factor, bifactor and bifactor S-1, which all had acceptable fit. We did not assess MI of unidimensional model due to its overall poor fit. We encountered some issues with fitting the bifactor model across all covariates apart from period (to be concise results are not presented).

Convergence issues were not found for the bifactor S-1 model, which is typically found to be more stable and more likely to converge than bifactor model (see Eid et al., 2017 for more details). This was another advantage of the bifactor S-1 model in the context of our study.

Both for 2-factor and bifactor S-1 models, we found evidence for scalar invariance across genders, and modes of collection as well as metric invariance across age groups and languages. There was an indication of scalar non-invariance for age-groups and languages. This was reflected by a substantially worse fit of the scalar models compared to the metric ones for bifactor and 2-factor models. Also, the overall fit of 2-factor scalar models was poor.

To better understand what items may have contributed to scalar non-invariance, we further tested invariance separately for the positive affect and life satisfaction factor and negative affect factor (see Table S2 in the Appendix). Within the positive affect and life satisfaction factor, we found a higher intercept for health satisfaction among young people (8.24) compared to middle-aged (7.73) and older individuals (7.62). This means that when comparing groups with the same level of positive affect and life satisfaction (equal zero in this case), young people, on average, tend to have a higher health satisfaction than other age groups. Importantly, health satisfaction appears equally important for positive affect and life satisfaction, as indicated by equal loadings across age groups in the metric model. The intercept of satisfaction with leisure was higher among older people (8.45) than middle-aged (7.66) and young people (7.48). Allowing the intercepts of these two items to vary across groups resulted in a significant improvement in the fit of the model.

We found a higher intercept for the worry item among the participants who completed their questionnaire in French (4.77) compared with Italian (3.88) and German (2.53). Allowing the intercept to vary across groups substantially improved the model fit. French speaking individuals scored on average 5.13 (sd = 2.34) on the worry item, substantially higher than Italian- (mean = 3.99, sd = 2.31) and German-speaking respondents (mean = 2.87, sd = 2.12).

Table 6. Assessment of measurement invariance across genders, age groups, period, languages and modes of questionnaire completion.

Model	Grouping variable	Invariance types	χ2	*df*	CFI	SRMR	RMSEA	Contrast	Δχ2	Δ*df*	p	ΔCFI	ΔSRMR	ΔRMSEA
Two-factor (correlated)	Gender	Configural	3505.25	68	0.919	0.051	0.081							
		Metric	3525.85	76	0.919	0.051	0.077	M vs C	20.60	8	0.008	0.000	0.000	-0.004
		Scalar	3714.52	84	0.914	0.052	0.075	S vs M	188.67	8	<0.001	-0.004	0.001	-0.002
	Age	Configural	3697.38	102	0.918	0.051	0.083							
		Metric	3876.23	118	0.914	0.055	0.079	M vs C	178.84	16	<0.001	-0.004	0.004	-0.004
		Scalar	5543.73	134	0.876	0.065	0.089	S vs M	1667.51	16	<0.001	-0.038	0.010	0.010
	Period	Configural	3114.05	68	0.914	0.049	0.081							
		Metric	3144.42	76	0.914	0.050	0.077	M vs C	30.37	8	<0.001	-0.001	0.001	-0.004
		Scalar	3526.78	84	0.903	0.053	0.077	S vs M	382.36	8	<0.001	-0.011	0.003	0.001
	Language	Configural	3697.41	102	0.916	0.053	0.083							
		Metric	3785.65	118	0.915	0.054	0.078	M vs C	88.24	16	<0.001	-0.002	0.001	-0.005
		Scalar	6416.09	134	0.854	0.071	0.096	S vs M	2630.44	16	<0.001	-0.061	0.017	0.018
	Mode	Configural	3435.42	68	0.920	0.050	0.080							
		Metric	3517.32	76	0.918	0.052	0.077	M vs C	81.89	8	<0.001	-0.002	0.002	-0.004
		Scalar	3846.54	84	0.911	0.054	0.077	S vs M	329.22	8	<0.001	-0.008	0.002	0.000
Bifactor S-1	Gender	Configural	1547.48	62	0.965	0.027	0.056							
		Metric	1588.82	74	0.964	0.028	0.052	M vs C	41.34	12	<0.001	-0.001	0.002	-0.004
		Scalar	1871.76	82	0.958	0.031	0.053	S vs M	282.93	8	<0.001	-0.006	0.003	0.002
	Age	Configural	1762.39	93	0.962	0.028	0.059							
		Metric	1910.85	117	0.959	0.034	0.055	M vs C	148.47	24	<0.001	-0.003	0.006	-0.004
		Scalar	3661.73	133	0.919	0.048	0.072	S vs M	1750.88	16	<0.001	-0.040	0.014	0.017
	Period	Configural	1497.51	62	0.960	0.028	0.058							
		Metric	1540.12	74	0.959	0.030	0.054	M vs C	42.61	12	<0.001	-0.001	0.002	-0.004
		Scalar	1931.33	82	0.948	0.034	0.057	S vs M	391.21	8	<0.001	-0.011	0.004	0.004
	Language	Configural	1710.33	93	0.962	0.029	0.058							
		Metric	1829.97	117	0.960	0.032	0.054	M vs C	119.64	24	<0.001	-0.002	0.003	-0.005
		Scalar	4328.28	133	0.903	0.050	0.079	S vs M	2498.32	16	<0.001	-0.058	0.018	0.025
	Mode	Configural	1580.86	62	0.964	0.028	0.057							
		Metric	1620.19	74	0.963	0.029	0.052	M vs C	39.33	12	<0.001	-0.001	0.001	-0.004
		Scalar	1962.42	82	0.955	0.032	0.055	S vs M	342.23	8	<0.001	-0.008	0.004	0.002

[1]Negative variance in women in energy and optimism item.
[2]Negative variance in age 14-35 in depression and anxiety item and in age 36-65 in energy and optimism item.
[3]Negative variance in Italian in energy and optimism item.

Source: SHP (2020) author's own calculations.

5. Discussion

In this study, we demonstrated that psychometrically robust wellbeing measures can be derived using individual items of the Swiss Household Panel, capturing two subcomponents: 1) positive affect and life satisfaction, and 2) negative affect. Its use across future studies, in addition to or instead of individual items, can help to improve comparability of research based on the SHP. Some caution, however, needs to be taken when an overall score of wellbeing is used, as we found evidence that it is unlikely to be unidimensional concept when operationalised with the SHP items. Also, results related to comparisons across age or lingual groups should be interpreted carefully, as we found evidence that these may be interpreted differently (they are non-invariant). Aggregated observed scores (e.g., by summing up the items) and factors scores were nearly perfectly correlated. They were also found to have highly comparable associations across basic sociodemographic groups. Hence, there appears to be little advantage of using more complex statistical techniques to obtain scores from latent factors.

5.1 Main findings and implications

Our study suggests that wellbeing as measured in the SHP is not a unidimensional construct. Therefore, we would recommend not using an overall wellbeing score. Our findings are consistent in several aspects with evidence from previous studies using a similar theoretical framework for the operationalisation of wellbeing. First, previous research also found that a unidimensional model, capturing overall wellbeing, tended to have a poor fit, with bifactor models performing best (Chen et al., 2013; Jovanović, 2015; Kaufman et al., 2022; Nima et al., 2020). Second, models including correlated subcomponents of wellbeing (i.e., positive affect, negative affect, life satisfaction), without a higher order factor, would typically have worse, but acceptable fit (Golino & Christensen, 2022). Overall, the conclusion from these studies is that both using general wellbeing as a higher order factor, or its components have their merits depending on the purpose of the analysis. The components of wellbeing tend to be moderately correlated, as shown here and by other studies, hence possibly having specific associations with exposure variables (Daniel-González et al., 2020).

One of the main findings of our study, which diverges from most of the existing literature, is that indicators of positive affect and life satisfaction were found to load on the same factor, rather than two separate (but positively correlated) factors (Chen et al., 2013; Diener et al., 2002; Kaufman et al., 2022; Nima et al., 2020). Previous studies varied in design, which may explain some of the differences found. For instance, response options typically varied in their range across positive affect and life satisfaction questions (Heinrich et al., 2023; Jovanović, 2015; Kaufman et al., 2022), where in the SHP they were harmonised (from 0 to 10), which may have inflated the correlation between items. If researchers may, of course, choose to treat these items as separate, if they have strong theoretical reasons for it. Statistical evidence obtained in our study suggests keeping these items combined.

As in previous studies, we found that bifactor and bifactor S-1 models fit the data particularly well (Jovanović, 2015; Kaufman et al., 2022). However, the choice of the model should not be determined solely by fit indices. Bifactor models nearly always produce a better fit than alternative specifications with lower order factors (Heinrich et al., 2023). The key reason why researchers would be interested in modelling wellbeing as a bifactor could be to obtain a “purer” form of the negative affect, without the shared variance with items capturing positive affect and life satisfaction. Using such a variable (i.e., as an exposure or outcome) should, however, be theoretically justified (Heinrich et al., 2023; Nima et al., 2020). Analyses including negative affect derived from a bifactor model are likely to produce different results than when using an aggregated score, as these do not correlate perfectly. This was the case in our study, where age appeared to have associations of opposite signs across both forms of negative affect when used as an outcome.

We found scalar invariance across most comparison groups for the wellbeing measure and its subcomponents, including period (2006, 2013, 2020), gender (men, women), mode of data collection (CATI, CAWI). Some caution should be taken when comparing positive affect and life satisfaction across age groups and negative affect across survey languages. This is not an uncommon finding. A systematic review of 27 studies conducted over three decades showed that scalar invariance across different age and cultural groups was rarely achieved with the Satisfaction with Life Scale (Emerson et al., 2017). In terms of positive affect, one ought to be cautious when analysing age differences in this wellbeing subcomponent, as young people typically have a higher health satisfaction than older people, when an overall level of positive affect and life satisfaction is accounted for. Similarly, older people have a higher level of satisfaction with their leisure activities. One may choose to remove these items to achieve partial invariance when age comparisons are of key interest or conduct sensitive analysis.

For negative affect, those who completed their questionnaire in French had on average a higher score on the worry item, compared with German and Italian speakers, when their levels of negative affect were kept equal. This is in line with another study, in which three items capturing sadness, fear, and anger were found to be culturally noninvariant, but not worry item (Jovanović et al., 2022). A potential reason for non-invariance in the SHP may be due to translating “worry” as “Angst” in German, which typically is understood as “fear”. “Worry” was translated as “inquiétude” in French and “inquietudine” in Italian. This, however, does not explain why the French speakers differed from the Italian speakers, which may be due to cultural differences.

5.2 Strengths and limitations

We demonstrate that from items that may not come from a single battery, but that cover the main domains of wellbeing, a good-quality composite measure can be derived. Moreover, we assessed psychometric properties of this measure, such as measurement invariance, for future researchers to know where caution is warranted when deciding on the research question or drawing conclusion from the results.

5.3 Conclusion

In this manuscript, we presented the efforts to develop a statistically robust measure of wellbeing in the SHP. This is to help researchers make more informed decisions about the choice of the wellbeing measure when using this dataset. A composite wellbeing variable may be desirable in certain contexts, for instance, when assessing trends in wellbeing over time or compare wellbeing across subpopulations. Deriving a measure from individuals items requires substantial amount of work, and this paper could improve efficiency in conducting research on wellbeing in SHP, as it includes detailed information on such attempts. We refrained ourselves from giving overly prescriptive recommendations, as the choice of the measure should be informed theoretically depending on the research question and statistical information should still be given by researchers as it is largely sample-specific. Providing statistical information about measurement choices can comparability and replicability of findings. It should also be kept in mind that there are some advantages of using individual items, as opposed to composites. Item-by-item analyses can provide important insights into specific causal mechanisms linking a given exposure with wellbeing or explaining how aspects of wellbeing translate into other outcomes. This can be particularly relevant when thinking of public health interventions. Our study still has the potential to serve as a useful guide when a researcher is interested in individual wellbeing items, as it provides information on their availability, distribution and relationship with other indicators of wellbeing in the SHP.

References

Barbuscia, A., & Comolli, C. (2021). Gender and socioeconomic inequalities in health and wellbeing across age in France and Switzerland. *Vienna Yearbook of Population Research*, *19*(1), 215-254. https://EconPapers.repec.org/RePEc:vid:yearbk:v:19:y:2021:i:1:oid:0x003c8cc6

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, *28*, 97-104. https://doi.org/10.1111/j.2044-8295.1937.tb00863.x

Beauducel, A., & Wittmann, W. W. (2005). Simulation Study on Fit Indexes in CFA Based on Data With Slightly Distorted Simple Structure. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(1), 41-75. https://doi.org/10.1207/s15328007sem1201_3

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588-606. https://doi.org/10.1037/0033-2909.88.3.588

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. The Guildford Press.

Burton-Jeangros, C., Duvoisin, A., Consoli, L., Fakhoury, J., & Jackson, Y. (2021). The well-being of newly regularized migrant workers: Determinants of their satisfaction with life as compared to undocumented migrant workers and regular local residents. *Comparative Migration Studies*, *9*(1), 42. https://doi.org/10.1186/s40878-021-00244-2

Busseri, M. A., & Sadava, S. W. (2011). A review of the tripartite structure of subjective well-being: implications for conceptualization, operationalization, analysis, and synthesis. *Pers Soc Psychol Rev*, *15*(3), 290-314. https://doi.org/10.1177/1088868310391271

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chen, F. F., Jing, Y., Hayes, A., & Lee, J. M. (2013). Two Concepts or Two Approaches? A Bifactor Analysis of Psychological and Subjective Well-Being. *Journal of Happiness Studies*, *14*(3), 1033-1068. https://doi.org/10.1007/s10902-012-9367-x

Chesters, J., Simona, J., & Suter, C. (2021). Cross-National Comparison of Age and Period Effects on Levels of Subjective Well-Being in Australia and Switzerland During Volatile Economic Times (2001–2016). *Social Indicators Research*, *154*(1), 361-391. https://doi.org/10.1007/s11205-020-02567-6

Christensen, A. P., & Golino, H. (2021). Estimating the Stability of Psychological Dimensions via Bootstrap Exploratory Graph Analysis: A Monte Carlo Simulation and Tutorial. *Psych*, *3*(3), 479-500. https://www.mdpi.com/2624-8611/3/3/32

Comolli, C. L., Bernardi, L., & Voorpostel, M. (2021). Joint Family and Work Trajectories and Multidimensional Wellbeing. *European Journal of Population*, *37*(3), 643-696. https://doi.org/10.1007/s10680-021-09583-3

Costantini, G., & Epskamp, S. (2017). *EstimateGroupNetwork: Perform the Joint Graphical Lasso and selects tuning parameters. R package version 0.1, 2.*

Daniel-González, L., de la Rubia, J. M., Valle, O. A., & García-Cadena, C. H. (2020). Structure analysis of subjective well-being. *Salud Mental.*, *43*(3), 119-127. DOI: 10.17711/SM.0185-3325.2020.017

Dawson-Townsend, K. (2019). Social participation patterns and their associations with health and well-being for older adults. *SSM Popul Health*, *8*, 100424. https://doi.org/10.1016/j.ssmph.2019.100424

Debnar, C., Carrard, V., Morselli, D., Michel, G., Bachmann, N., & Peter, C. (2021). The longitudinal impact of a chronic physical health condition on subjective well-being. *Health Psychol*, *40*(6), 357-367. https://doi.org/10.1037/hea0001090

Diener, E. (1984). Subjective well-being. *Psychol Bull*, *95*(3), 542-575. https://www.ncbi.nlm.nih.gov/pubmed/6399758

Diener, E., Lucas, R. E., & Oishi, S. (2002). Subjective well-being: The science of happiness and life satisfaction. In C. R. S. S. J. Lopez (Ed.), *Handbook of positive psychology* (pp. 63–73). Oxford University Press.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychol Methods*, *22*(3), 541-562. https://doi.org/10.1037/met0000083

Emerson, S. D., Guhn, M., & Gadermann, A. M. (2017). Measurement invariance of the Satisfaction with Life Scale: reviewing three decades of research. *Quality of Life Research*, *26*(9), 2251-2264. http://www.jstor.org/stable/44856122
Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*(4), 1–18. https://doi.org/10.18637/jss.v048.i04
Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychol Methods*, *23*(4), 617-634. https://doi.org/10.1037/met0000167
Gilman, R., Huebner, E. S., & Laughlin, J. E. (2000). A First Study of the Multidimensional Students' Life Satisfaction Scale with Adolescents. *Social Indicators Research*, *52*(2), 135-160. https://doi.org/10.1023/A:1007059227507
Golino, H., & Christensen, A. P. (2022). *EGAnet: Exploratory Graph Analysis – A framework for estimating the number of dimensions in multivariate data using network psychometrics. R package version 1.1.1.*
Gustafsson, J.-E., & Åberg-Bengtsson, L. (2010). *Unidimensionality and interpretability of psychological instruments. In S. E. Embretson (Ed.), Measuring psychological constructs: Advances in model-based approaches* (Measuring psychological constructs: Advances in model-based approaches, Issue. A. P. Association. https://doi.org/10.1037/12074-005
Heinrich, M., Geiser, C., Zagorscak, P., Burns, G. L., Bohn, J., Becker, S. P., Eid, M., Beauchaine, T. P., & Knaevelsrud, C. (2023). On the Meaning of the "P Factor" in Symmetrical Bifactor Models of Psychopathology: Recommendations for Future Research From the Bifactor-(S–1) Perspective. *Assessment*, *30*(3), 487-507. https://doi.org/10.1177/10731911211060298
Henning, G., Baumann, I., & Huxhold, O. (2023). Historical and Cross-Country Differences in Life Satisfaction Across Retirement in Germany and Switzerland From 2000 to 2019. *The Journals of Gerontology: Series B*. https://doi.org/10.1093/geronb/gbad066
Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika 2*, 41–54. https://doi.org/10.1007/bf02287965
Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling. R package version 0.5-6. Retrieved from https://CRAN.R-
Jovanović, V. (2015). A bifactor model of subjective well-being: A re-examination of the structure of subjective well-being. *Personality and Individual Differences*, *87*, 45-49. https://doi.org/https://doi.org/10.1016/j.paid.2015.07.026
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. https://doi.org/10.1177/10731911211021494
Kaufman, V. A., Horton, C., Walsh, L. C., & Rodriguez, A. (2022). The Unity of Well-Being: An Inquiry into the Structure of Subjective Well-Being Using the Bifactor Model. *International Journal of Applied Positive Psychology*, *7*(3), 461-486. https://doi.org/10.1007/s41042-022-00077-z
Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1
Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology*, *71*(3), 616-628. https://doi.org/10.1037/0022-3514.71.3.616
Luhmann, M. (2017). The development of subjective well-being. In J. Specht (Ed.), Personality development across the lifespan (pp. 197–218). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-804674-6.00013-2
Mansolf, M., & Reise, S. P. (2016). Exploratory Bifactor Analysis: The Schmid-Leiman Orthogonalization and Jennrich-Bentler Analytic Rotations. *Multivariate Behav Res*, *51*(5), 698-717. https://doi.org/10.1080/00273171.2016.1215898
McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287-2305. https://doi.org/10.3758/s13428-020-01398-0
Metler, S. J., & Busseri, M. A. (2017). Further Evaluation of the Tripartite Structure of Subjective Well-Being: Evidence From Longitudinal and Experimental Studies. *J Pers*, *85*(2), 192-206. https://doi.org/10.1111/jopy.12233

Nima, A. A., Cloninger, K. M., Lucchese, F., Sikström, S., & Garcia, D. (2020). Validation of a general subjective well-being factor using Classical Test Theory. *PeerJ*, *8*, e9193. https://doi.org/10.7717/peerj.9193

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10.2*, 191-218. https://www-complexnetworks.lip6.fr/~latapy/Publis/communities.pdf

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1 - 36. https://doi.org/10.18637/jss.v048.i02

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research*, *8*(2), 23-74. https://psycnet.apa.org/record/2003-08119-003

Schröder, M. (2020). Men Lose Life Satisfaction with Fewer Hours in Employment: Mothers Do Not Profit from Longer Employment—Evidence from Eight Panels. *Social Indicators Research*, *152*(1), 317-334. https://doi.org/10.1007/s11205-020-02433-5

SHP Group, Living in Switzerland Waves 1-23 + Covid 19 data. (2023). FORS - Swiss Centre of Expertise in the Social Sciences. Financed by the Swiss National Science Foundation, distributed by FORS, Lausanne. DOI: https://doi.org/10.48573/1nav-wy98

Team, R. C. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.R-project.org/

Tillmann, R., Voorpostel, M., Antal, E., Dasoki, N., Klaas, H., Kuhn, U., Lebert, F., Monsch, G.-A., & Ryser, V.-A. (2021). The Swiss Household Panel (SHP). Jahrbücher für Nationalökonomie und Statistik, 242(3), 403-420. https://doi.org/10.1515/jbnst-2021-0039

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(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Węziak-Białowolska, D. (2016). Attendance of cultural events and involvement with the arts—impact evaluation on health and well-being from a Swiss household panel survey. *Public Health*, *139*, 161-169. https://doi.org/10.1016/j.puhe.2016.06.028 https://doi.org/https://doi.org/10.1016/j.puhe.2016.06.028

Widaman, K. F., & Revelle, W. (2023). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *55*(2), 788-806. https://doi.org/10.3758/s13428-022-01849-w

Zhang, Z., & Yuan, K. H. (2016). Robust Coefficients Alpha and Omega and Confidence Intervals With Outlying Observations and Missing Data: Methods and Software. *Educ Psychol Meas*, *76*(3), 387-411. https://doi.org/10.1177/0013164415594658

Appendix

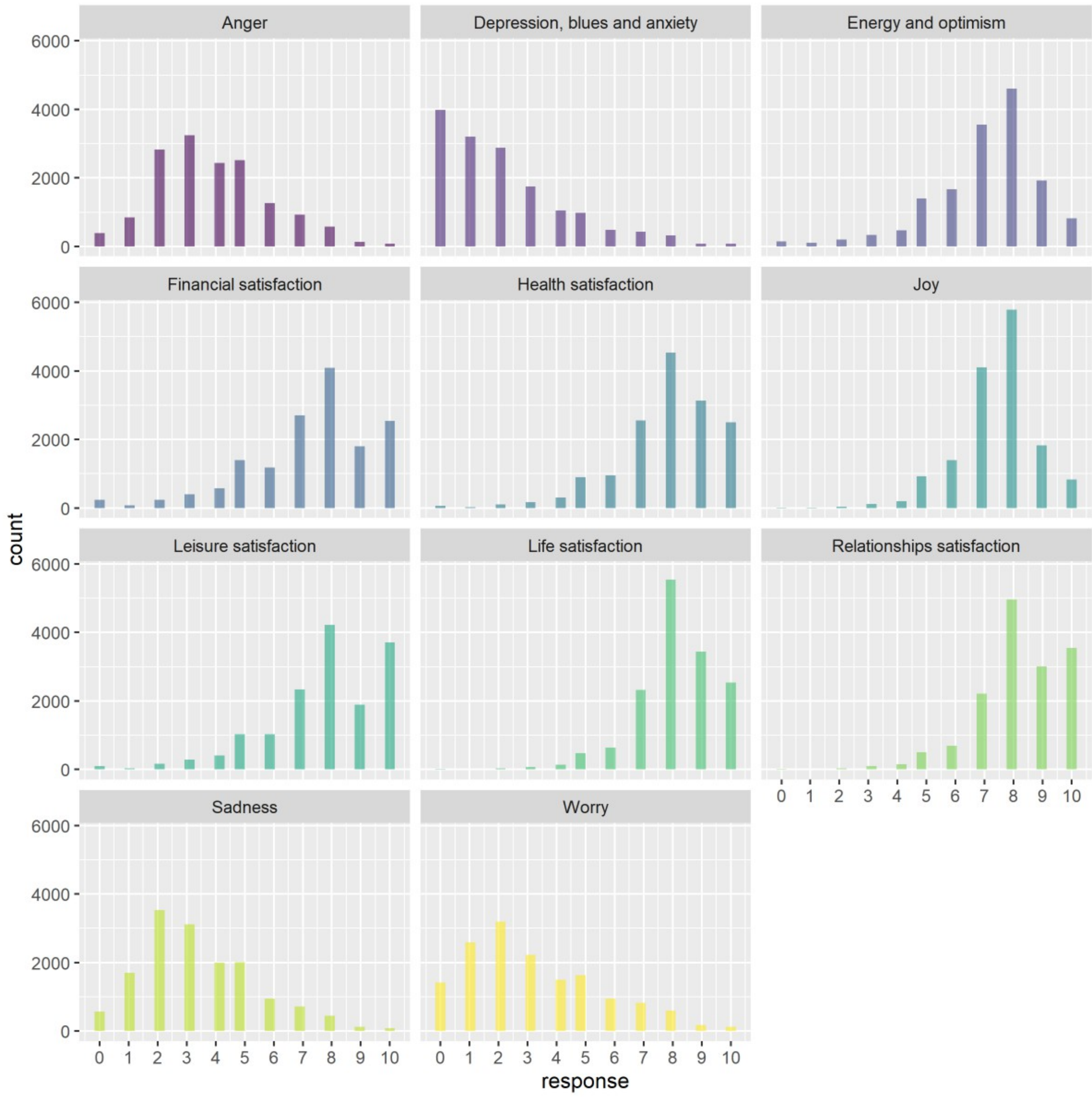


Figure S1. Distribution of indicators of wellbeing.

Source: SHP (2020) author's own calculations.

Table S1. Factor loadings of all items across four factor models.

	Unidimensional	2-factor		Bifactor			Bifactor S-1	
	Wellbeing	PALS	Negative affect	Wellbeing	PALS	Negative affect	Wellbeing	Negative affect
Life satisfaction	0.68	0.76		0.76	0.23		0.75	
Joy	0.57	0.64		0.63	0.04		0.65	
Health satisfaction	0.50	0.53		0.52	0.02		0.53	
Energy and optimism	0.57	0.58		0.72	-0.70		0.59	
Relationships satisfaction	0.53	0.59		0.58	0.15		0.59	
Leisure time satisfaction	0.44	0.50		0.48	0.11		0.49	
Anger	-0.42		0.53	-0.27		0.47	-0.27	0.47
Sadness	-0.61		0.80	-0.40		0.74	-0.41	0.74
Worry	-0.52		0.69	-0.31		0.63	-0.32	0.62
Depression and anxiety	-0.72		0.70	-0.59		0.41	-0.61	0.40

Source: SHP (2020) author’s own calculations.

Table S2. Assessment of measurement invariance for positive affect and life satisfaction and negative affect across genders, age groups, period, languages and modes of questionnaire completion.

Model	Grouping variable	Invariance types	χ2	*df*	CFI	SRMR	RMSEA	Contrast	Δχ2	Δ*df*	p	ΔCFI	ΔSRMR	ΔRMSEA
Positive only	Gender	Configural	712.33	18	0.966	0.027	0.071							
		Metric	729.70	23	0.965	0.028	0.063	M vs C	17.38	5	0.004	-0.001	0.001	-0.008
		Scalar	803.83	28	0.961	0.030	0.060	S vs M	74.13	5	<0.001	-0.003	0.002	-0.003
	Age	Configural	823.48	27	0.961	0.028	0.076							
		Metric	874.59	37	0.959	0.033	0.066	M vs C	51.11	10	<0.001	-0.002	0.004	-0.009
		Scalar	2300.84	47	0.891	0.058	0.097	S vs M	1426.25	10	<0.001	-0.069	0.025	0.030
	Period	Configural	620.92	18	0.965	0.027	0.070							
		Metric	640.93	23	0.964	0.029	0.063	M vs C	20.01	5	0.001	-0.001	0.002	-0.007
		Scalar	937.92	28	0.946	0.037	0.069	S vs M	297.00	5	<0.001	-0.017	0.008	0.006
	Language	Configural	739.24	27	0.965	0.027	0.072							
		Metric	789.62	37	0.963	0.031	0.063	M vs C	50.38	10	<0.001	-0.002	0.003	-0.009
		Scalar	902.31	47	0.957	0.033	0.060	S vs M	112.69	10	<0.001	-0.005	0.002	-0.003
	Mode	Configural	684.29	18	0.966	0.026	0.069							
		Metric	699.13	23	0.966	0.027	0.062	M vs C	14.84	5	0.011	0.000	0.001	-0.008
		Scalar	1016.52	28	0.950	0.035	0.068	S vs M	317.39	5	<0.001	-0.016	0.007	0.006
Negative only	Gender	Configural	152.85	4	0.991	0.016	0.070							
		Metric	164.97	7	0.990	0.018	0.054	M vs C	12.12	3	0.007	-0.001	0.002	-0.015
		Scalar	289.98	10	0.983	0.026	0.060	S vs M	125.01	3	<0.001	-0.008	0.008	0.006
	Age	Configural	181.47	6	0.990	0.016	0.076							
		Metric	264.57	12	0.985	0.027	0.064	M vs C	83.10	6	<0.001	-0.005	0.010	-0.011
		Scalar	524.18	18	0.970	0.036	0.074	S vs M	259.61	6	<0.001	-0.015	0.010	0.010
	Period	Configural	115.42	4	0.991	0.015	0.064							
		Metric	125.49	7	0.991	0.017	0.050	M vs C	10.08	3	0.018	-0.001	0.003	-0.014
		Scalar	203.65	10	0.985	0.023	0.053	S vs M	78.16	3	<0.001	-0.006	0.006	0.003
	Language	Configural	210.66	6	0.988	0.018	0.082							
		Metric	258.16	12	0.985	0.024	0.063	M vs C	47.50	6	<0.001	-0.002	0.007	-0.018
		Scalar	2801.15	18	0.835	0.087	0.174	S vs M	2542.99	6	<0.001	-0.150	0.063	0.110
	Mode	Configural	159.15	4	0.991	0.016	0.071							
		Metric	185.72	7	0.989	0.020	0.058	M vs C	26.57	3	0.000	-0.001	0.005	-0.013
		Scalar	202.77	10	0.988	0.021	0.050	S vs M	17.05	3	0.001	-0.001	0.001	-0.008

[1] Negative variance in women in energy and optimism item.
[2] Negative variance in age 14-35 in depression and anxiety item and in age 36-65 in energy and optimism item.
[3] Negative variance in Italian in energy and optimism item.

Source: SHP (2020) author's own calculations.