ARTICLE

Are Danes really happier than Swedes and Australians? A comparative analysis of subjective well-being measurement issues in Denmark, Sweden, and Australia

Filip Fors Connolly · Johannes Frech · Bengt Brülde · Christopher Kullenberg

Abstract: This study examines the reliability of cross-national subjective well-being (SWB) survey data, by corroborating the persistently high SWB-rankings of Denmark through a comparison to Sweden and Australia, two countries with similar or superior social, economic, and health indicators. Our research focuses on both the affective and the cognitive component of SWB, with a special emphasis on the affective component. We investigate four potential measurement issues that could contribute to Denmark's elevated SWB scores: linguistic inconsistencies in survey translations, variations in answering scale usage, recall bias of affect, and positivity bias in life satisfaction judgments. To address these concerns, we utilize multi-group confirmatory factor analysis, analyse emotion-focused anchoring vignettes, examine affect items across varying time frames, and contrast overall life satisfaction assessments with those of domain-specific satisfaction. Despite accounting for various potential measurement issues, our results reveal that Danes consistently report higher SWB than their Swedish and Australian counterparts, although the differences are small for several of the measures. This finding implies that the SWB survey-data is reliable in this case, and that Denmark's high SWB rankings are not attributable to measurement biases but may indeed signify genuinely high levels of SWB. This paper adds to the growing body of literature on cross-national SWB comparisons and might provide insights for researchers aiming to compare well-being across countries.

Keywords: country-comparison; subjective well-being; affective well-being; life satisfaction; measurement equivalence

1. Introduction

Subjective well-being (SWB) comprises people's satisfaction with their life and experiences of positive and negative affect (affective well-being, AWB). The relevance of SWB is self-evident from both a societal and an individual perspective since SWB is an outcome that is good in itself rather than a means to something else. SWB clearly differs from other outcomes such as income, that is primarily a means to an end, rather than being a constituent part of the good life. In line with this idea, SWB constitutes a key aspect of quality of life according to both philosophers (Brülde, 2007) and psychologists (Diener, 2009). Organizations such as the United Nations (Helliwell et al., 2017) and the OECD (2011) recommend using measures of SWB when assessing national levels of welfare. Although research has shed light on many factors that may contribute to SWB across nations (cf. Suh & Koo, 2008), our current understanding is limited by an overemphasis on life satisfaction (LS) compared to AWB, and a lack of methodological studies to

Filip Fors Connolly
Umeå University, Sweden
filip.fors@umu.se

validate existing measures of SWB. However, the usefulness of SWB survey data for assessments and policy is contingent on the reliability of the data.

Numerous survey studies have underscored Denmark's status as possibly the world's happiest country across both dimensions of SWB (Helliwell et al., 2017). While Finland has recently edged slightly ahead in terms of LS according to the United Nations' global SWB rankings (Helliwell et al., 2023), Denmark is still rated as second and has the highest average score in the last decades. It can also be noted that Denmark surpasses Finland in positive affect in the same study (ibid.).

However, despite the recurrent Danish lead in SWB based on international surveys, the accuracy of the data can be doubted. First, it is not obvious why Denmark would stand out from several other countries that show similar (or sometimes better) achievements on established cross-national determinants of SWB. For instance, when looking at United Nations ranking of human development across nations (Jaha et al., 2017), countries like Australia and Germany score as well or better while scoring clearly lower than Denmark on measures of SWB (Helliwell et al., 2017). Furthermore, Sweden, a country with striking similarities to Denmark regarding social, economic, cultural and institutional factors also score significantly lower on SWB[1], despite having higher life expectancy (Knudsen, 2019). Although cultural factors like "hygge" or "frisind"[2], as opposed to e.g. economic or health factors, may account for the elevated levels of SWB in Denmark, the specific influences underlying these disparities still need to be determined. This raises the question: could the high levels of reported SWB in Denmark be, at least in part, attributed to measurement issues? At least four measurement challenges could contribute to potential discrepancies in cross-national comparisons of SWB, where Denmark serves as an intriguing case for exploring these issues given its reputation as one of the world's happiest, if not *the* happiest, country.

Firstly, linguistic discrepancies in survey question translations may result in non-equivalent interpretations of SWB items across countries. For instance, in the case of Denmark, Lolle & Andersen (2016) found that translations of adjectives such as "satisfied" or "happy" into Danish make response comparability between Denmark and other countries more difficult. Secondly, there is evidence to suggest that Danish respondents might employ answering scales differently, potentially overreporting on the positive end of the scale, which Angelini et al. (2014) found support for, using anchoring vignettes. Importantly, when adjusting the mean values for different countries based on the vignettes, Swedes reported higher LS than Danes while the opposite was true without correcting for the vignettes. Lastly, biased recall of emotions as well as positivity bias, potentially stemming from Denmark's cultural emphasis on contentment and social harmony, may prompt individuals to recall and report SWB levels that align with these values (Oishi, 2018). This could result in overreported SWB levels compared to other countries, independent of the previously mentioned concerns.

To address these concerns, our study employs a comprehensive methodology with four key components. We begin by using Multi-Group Confirmatory Factor Analysis to assess measurement invariance of SWB constructs across countries, ensuring observed differences are not due to linguistic issues. Next, we utilize emotion-focused anchoring vignettes to control for potential differences in scale usage between countries. We then analyze variations in affect reports across different time frames to account for potential recall bias. Finally, we examine both global and domain-specific measures of LS to address possible positivity bias.

---

[1] In the last World Happiness Report, the average life satisfaction in Denmark was 7.6 and in Sweden it was 7.4.
[2] "Hygge" is a Danish concept that embodies the feeling of cozy contentment through enjoying simple pleasures in life. "Frisind" refers to an open-minded attitude towards diverse ideas and viewpoints.

Our analysis focuses on contrasting Denmark with Sweden and Australia. The selection of these countries is partly based on the availability of comprehensive data.[3] However, more importantly, Sweden and Australia serve as compelling comparison points because they share similar societal and economic conditions with Denmark, which are relevant to SWB. Despite these similarities, Sweden and Australia exhibit lower SWB scores compared to Denmark. This significant discrepancy raises concerns that measurement issues may be contributing to the high SWB scores reported in Denmark. By applying the aforementioned methodological approaches, we aim to provide a more accurate understanding of the validity of SWB measurements in these countries and shed light on the factors underlying the observed differences in reported well-being. In addition, unlike previous research, we emphasize AWB alongside LS due to its normative importance (Brülde, 2007) and its role in interpreting cross-country differences in LS (Kahneman & Riis, 2005).

1.1 Previous research

Most cross-national studies on SWB use single-item measures of LS or happiness without adjusting for reliability and validity. While studies on AWB often use multiple items, they are less common (Diener et al., 2010). Consequently, the reliability of cross-country SWB measurements remains uncertain. Denmark consistently scores high in these studies, which may be due to factors like economic wealth, life expectancy, trust, social cohesion, and institutional efficacy (Biswas-Diener et al., 2010; Helliwell et al., 2017). However, it's puzzling why Danes report higher SWB levels than residents of countries like Sweden and Australia, which match or surpass Denmark on several of these factors. This highlights the need to test for measurement issues to ensure accurate cross-country SWB comparisons.

1.1.1 Measurement issues due to linguistic factors

Measurement inequivalence due to linguistic problems is an important issue in cross-national survey research because it can be challenging to make accurate translations of survey items across different cultures and languages (Billiet, 2003; Saris & Gallhofer, 2007). As an illustration of this issue, Lolle & Andersen (2016) conducted an experiment comparing how Danish students answered single-item questions on happiness and LS in both English and Danish. They found that the answers differed significantly depending on the language used. The students reported higher levels of LS when they responded to the Danish questionnaire, while the reverse was the case for happiness. In addition, the authors also note that the translation of the word "extremely," which is used to define the end poles when assessing LS and happiness in the European Social Survey (ESS), has been translated to "særdeles" in the Danish questionnaire, a weaker term often translated to English as "highly" or "very." Thus, linguistic problems in survey questions as well as answering scales may partly be responsible for the high reported SWB in Denmark, not least when it comes to explaining Denmark's particularly high SWB in the ESS.

Multi-Group Confirmatory Factor Analysis (MGCFA) is a formal approach to address linguistic problems in cross-national research. MGCFA is a statistical technique that allows researchers to assess whether survey items function similarly across different languages by comparing the factor structure and measurement properties of a survey inventory. By applying MGCFA, researchers can determine if the same construct is being measured across different

---

[3] Comparing Denmark and Sweden exclusively could seem more intuitive due to their cultural and institutional parallels. However, given Australia's inclusion, a nation that excels in various cross-national determinants of SWB but records lower SWB levels than Denmark, omitting Australia from our analysis would represent a missed opportunity.

languages and identify translation or linguistic issues that may affect the measurement equivalence of the survey.

Several studies have employed MGCFA to investigate the measurement invariance of SWB across different countries and cultures. Jang et al. (2017) found that the factor structure of LS was invariant across 26 countries, suggesting that LS can be compared across many different cultures and languages. Similarly, Jovanović et al. (2022) found support for partial measurement invariance across 24 countries when assessing LS among adolescents. However, it is worth noting that Denmark was not included in either of these studies.

The MGCFA framework has also been used to explore differences in AWB between European nations, specifically in terms of Positive and Negative Affect derived from the CES-D 8 depression scale. Studies by Fors & Kulin (2016), Raudenská (2020), and Sortheix & Weber (2023) have shown partial measurement invariance between most European nations when comparing levels of PA and NA. Interestingly, adjusting for measurement equivalence using MGCFA affected country rankings in some cases. For example, Sortheix & Weber (2023) revealed that Denmark's ranking experienced a significant decline when latent error-corrected measurements of positive affect were employed, as opposed to using manifest measurements. This finding suggests that potential translation issues may have influenced the results, highlighting the importance of addressing linguistic problems in cross-national research.

While these studies provide valuable insights into the measurement invariance of SWB across countries and cultures, they have some limitations. The findings from Sortheix & Weber's 2023 study, which employed latent affect measures derived from items intended to assess depression, in conjunction with Lolle & Andersen's 2016 small-scale study on students using single-item measures of LS and happiness, suggest that translation issues may have a partial influence on the high levels of SWB reported in Denmark. However, the constrained focus of Lolle & Andersen's research on a student cohort and their reliance on single-item measures, along with Sortheix & Weber's utilization of depression-focused items, render it challenging to draw definitive conclusions from these studies alone. To address these limitations and provide a more comprehensive assessment of language-related issues in SWB measurement, the current study employs MGCFA on a comprehensive set of items specifically designed to measure both components of SWB.

1.1.2 Measurement issues due to answering scales

Setting linguistic issues in survey questions aside, scholars have noted that respondents may use response scales differently across countries. To address this in cross-group comparisons, anchoring vignettes have been proposed as a promising method (Salomon et al., 2004). These vignettes involve brief descriptions of hypothetical individuals, which respondents evaluate using the same scale they use for their own situation (King & Wand, 2007). If systematic differences in ratings emerge at the group level, this information can be used to calibrate respondents' self-assessments.

Anchoring vignettes were applied to SWB by Angelini et al. (2014), who suggested Denmark's high LS could be caused by Danes interpreting the answering scale differently. They analyzed data from ten European countries using vignette methodology and found that without accounting for differences in vignette evaluations, Denmark ranked highest in LS. However, when these differences were considered, rankings shifted significantly. For example, Sweden exhibited lower LS scores compared to Denmark in the uncorrected estimates, but when adjustments were made by using anchoring vignettes, the pattern was reversed, and Sweden scored higher than Denmark. However, it is unclear if the authors successfully accounted for

differences in scale interpretation or if the responses merely reflected participants' beliefs about how life circumstances influence LS. In the Danish culture, there's a prominent philosophy of appreciating and finding contentment in the present, encapsulated by the concept of 'hygge' (Bille 2015). This cultural ideal promotes the idea that happiness is to some extent independent of one's objective living conditions. Thus, the higher Danish vignette ratings observed in the study might reflect this cultural belief rather than a response bias.

Given the potential limitations of vignettes that may be conflated with interpretations of living conditions rather than scale use, we instead use affect-focused vignettes in this study. These vignettes are conceptually more aligned with SWB, emphasizing hypothetical individuals' emotional experiences, avoiding confusion between scale usage and interpretation of living conditions. By concentrating on emotional experiences directly related to SWB, we aim to minimize potential sources of bias and allow for clearer distinctions between cross-country differences in scale usage and genuine variations in assessments of well-being.

It's important to recognize that while the first measurement approach described earlier, MGCFA, assesses measurement invariance based on the relationships between items measuring SWB, it doesn't directly address how respondents interpret and use response scales in general. Consequently, even if strict invariance is established through MGCFA, systematic differences in reporting across groups may persist due to varying norms of scale use. In contrast, the vignette approach more directly tests for differences in scale use.

1.1.3 Measurement issues due to recall bias

Another important factor that can affect the accuracy of SWB measurements is cultural differences in how individuals remember and report their feelings, which may lead to an overestimation in self-reports of SWB in some countries (like Denmark). In previous studies (Oishi, 2010), this is referred to as recall bias. According to Oishi, cultural variations can impact how people evaluate and interpret their emotions, which may be related to differences in the significance of emotions across cultures. For example, some cultures might place a greater emphasis on expressing positive emotions and suppressing negative emotions, while others may encourage a more balanced expression of both positive and negative emotions (ibid.). Such cultural norms can influence not only how people genuinely experience their emotions from moment to moment but also how they remember them.

To combat recall biases, it is generally recommended to use methods that capture people's feelings in real-time, such as experience sampling (Csikszentmihalyi & Larson, 2014) or the day reconstruction method (Kahneman et al., 2004), instead of relying on retrospective ratings of affect or global evaluations (Kahneman & Riis, 2005; Oishi, 2010). This is because cultural expectations may contaminate retrospective measures to a higher extent, making it more difficult to accurately capture individuals' actual emotional experiences. By using experience sampling or similar techniques, researchers can more reliably capture people's moment-to-moment feelings, while minimizing the influence of recall biases.

Although previous studies have often compared LS to momentary affect rather than recalled affect, studies show that LS judgments are influenced by both an individual's present emotions and their recollection of past emotions (c.f. Wirtz et al., 2003). Oishi & Diener (2008) examined differences in LS between Japanese and American students and found that American students reported higher levels of LS, consistent with prior studies. In addition to LS, respondents were also asked to rate their current mood, but no significant differences were found between the two groups in this respect.

Further, using the day reconstruction method, a technique designed to emulate experience sampling, Kahneman et al. (2010) compared AWB of women in USA and France respectively, over the course of a day. Although American women reported slightly higher levels of LS, French women reported slightly higher levels of AWB throughout the day. Yet, no studies have systematically evaluated the relationship between different time frames and measures of AWB across countries, including comparisons with Denmark. We aim to address this gap by examining the relationship between reported affect ratings for the past week and reported affect ratings for the current day. Our rationale is that recall biases should more strongly influence affect reported for the past week compared to affect reported for the current day.

1.1.4 Measurement issues due to positivity bias

In addition to biases in memory recall, potential positivity biases can affect LS judgments through a desire to maintain a positive self-image. Global assessments of LS are often ambiguous, which allows individuals to incorporate social norms into their evaluation process. This phenomenon has been noted by Oishi (2010), who suggests that individuals may use this flexibility to maintain positive self-evaluations. However, when evaluating particular or more tangible areas of their life, individuals focus more on emotions and thoughts about the actual domains, which may reduce the relevant cultural influences. In support, Diener et al. (2000) found that people in countries that score high on overall LS measures differed more in their assessments of more global life domains (e.g. education) than more narrow domains (e.g. professors, textbooks, and lectures). In the current study, we aim to expand the understanding of this phenomenon by comparing global LS measures with domain-specific satisfaction measures. While our domain satisfaction measures remain somewhat broad (e.g., "Leisure"), we anticipate that global LS reports will be more significantly influenced by cultural norms than domain satisfaction reports.

While scale use differences, recall bias, and positivity bias may be influenced by common factors like cultural norms, we argue that treating them as distinct phenomena is warranted. Scale use differences specifically arise from varying interpretations of the response scale itself. In contrast, recall bias and positivity bias may stem from how cultural expectations shape the way individuals remember emotions and make judgments about their life.

2. Data and methods

2.1 Data

The dataset employed in this study was assembled during a research project focused on comparing SWB across various countries. Respondents were gathered via Qualtrics' web-survey panels during the fall of 2016 (http://www.qualtrics.com, Boas et al., 2018; Heen et al., 2014). The total sample size for this study consisted of 3,780 respondents. Participants completed an online questionnaire comprising 109 items. The questionnaire was provided in English for the Australian sample, in Danish for the Danish sample, and in Swedish for the Swedish sample. In this paper, we only analyze items related to SWB. The sociodemographic characteristics were largely similar across countries (see Appendix Table A1) and were largely representative of each population in terms of age and sex (see Appendix Table A2).

2.2 Measures

Item descriptions, abbreviations, and their respective answering scales can be found in Table 1 (below). To estimate latent country means and investigate potential linguistic measurement issues, six items covering reports of various emotions experienced during the past week were

used. Participants were requested to rate the frequency of experiencing three positive emotions—engagement, happiness, and relaxation—and three negative emotions—worry, sadness, and boredom—within the defined time frame. The terms employed to describe these emotions originated from the Swedish Core Affect Scale (SCAS, Västfjäll et al., 2007). To capture the frequency of these emotional experiences, we use response options, drawn from the Scale of Positive and Negative Experience (SPANE, Diener et al., 2009) which ranged from "Very Rarely or Never," through "Rarely" and "Sometimes," to "Often" or "Very Often or Always." Henceforth we refer to these measures as PA and NA.

Table 1. *Summary of survey items and corresponding answering scales*

Construct	Survey question/statement	Answering Scale
Positive and Negative Affect (6 items)	Please think about what you have been doing and experiencing during the past week. Then report how much you experienced each of the following feelings, using the scale below. For each item, select a number from 1 to 5. Engaged Happy Relaxed Worried Sad Bored	Very Rarely or Never (1) Rarely (2) Sometimes (3) Often (4) Very Often or Always (5)
Bipolar Mood (2 items)	How did you feel last week? (Mood Week) How do you feel today? (Mood Today)	In a very bad mood (0) In a very good mood (6)
Satisfaction With Life Scale (SWLS: 3 items)	In most ways, my life is close to my ideal. (SWLS1) The conditions of my life are excellent. (SWLS2) I am satisfied with my life. (SWLS3)	Strongly disagree (1) Disagree (2) Neither agree nor disagree (3) Agree (4) Strongly agree (5)
Global Life Satisfaction (GLS 1 item)	All things considered, how satisfied are you with your life as a whole nowadays? (GLS)	Extremely dissatisfied (0) Extremely satisfied (6)

To compare vignette ratings with respondents' self-reported AWB as well as to examine potential recall biases, two items were used to assess the overall quality of respondents' feelings during "the past week" and "today." Respondents were asked the questions, "How did you feel last week?" and "How do you feel today?" using a bipolar scale with endpoints labeled "In a very bad mood" and "In a very good mood." The response scales were identical to those employed in the vignettes described below. The rationale for selecting the time frame "the past week" was to maintain consistency with the SCAS items and numerous previous studies that employed retrospective reports of AWB (e.g., Fors & Kulin, 2016). In contrast, the "today" time frame aimed to minimize recall biases by focusing on respondents' current feelings. Henceforth we refer to these measures as "Bipolar Mood".

The evaluation of LS was carried out by utilizing three statements from the Satisfaction With Life Scale (Diener et al., 1985; Pavot & Diener, 2008). The participants were asked to express their level of agreement with the statements: "In most ways my life is close to my ideal," "The conditions of my life are excellent," and "I am satisfied with my life." The scale used for rating their agreement ranged from 1, indicating "Strongly disagree," to 5, signifying "Strongly agree."

Furthermore, we incorporated an additional single-item question taken from the European Social Survey, which asked: "All things considered, how satisfied are you with your life as a whole nowadays?" For this question, participants used a scale from 0 to 6, with "Extremely dissatisfied" and "Extremely satisfied" marking the extreme points of the scale. Henceforth we refer to the first LS measure as SWLS and the second LS measure as "Global LS".

Domain satisfaction was assessed using six items focusing on specific aspects of an individual's life. Respondents were asked to indicate their satisfaction with each domain using a response scale ranging from 0 ("Extremely dissatisfied") to 6 ("Extremely satisfied"). The items addressed satisfaction with one's "family life or close relationships", "financial situation", "looks", "leisure", "friends" and finally "work or studies".

2.2.1 Anchoring vignettes

To investigate whether the people in the three countries use the scales differently when assessing their AWB three different anchoring vignettes were used in the study. These vignettes covered three different scenarios with respect to affective wellbeing. Instead of describing the life circumstances of these individuals the focus was on describing their affective state. The first scenario illustrated a person with low levels of AWB, the second scenario described a person whose AWB was in between high and low, and the last scenario exemplified a person with high levels of AWB. In relation to each scenario respondents were asked to rate how good (or bad) the person in the scenario felt overall during the past four weeks by using a bipolar scale ranging from 0 to 6. The endpoints of the scale were labelled "In a very bad mood" and "In a very good mood" respectively.

Anchoring vignettes used in the present study:

Vignette 1: Anna feels depressed most of the time. She weeps frequently and feels hopeless about the future. She feels that she has become a burden on others and that she would be better dead. Overall, how did Anna feel during the last four weeks?

Vignette 2: Peter feels nervous and anxious but feels better in the company of people or when doing something that really interests him. When he is alone, he tends to feel useless and empty. Overall, how did Peter feel during the last four weeks?

Vignette 3: Lisa often feels engaged and interested in life. She enjoys her work and family life very much. When she wakes up in the morning, she almost always feels optimistic about the upcoming day. Overall, how did Lisa feel during the last four weeks?

2.3 Analytical strategy for examining cross-country measurement issues

All statistical analyses were conducted with R (Version 4.3.1; R Core Team, 2023). To investigate the levels of SWB in the three countries, we begin by analyzing the unadjusted values and comparing levels of SWB using three different measures of AWB and two measures of LS. This initial comparison shows whether our data corresponds to the data used in previous research, which typically show higher levels of SWB in Denmark compared to Sweden and Australia.

To investigate potential linguistic measurement issues, we conduct a Multi-Group Confirmatory Factor Analysis (MGCFA) using two distinct theoretical approaches to assess AWB. In the first model, we evaluate AWB as comprising two factors: Positive Affect and Negative Affect, using all six items taken from the SCAS (three items for each, NA and PA respectively). In the second model, we estimate a single AWB factor, utilizing SCAS items that capture pure valence by opposite adjectives (sad and happy) in combination with the bipolar

mood measure. This approach allows us to compare AWB as constituted of two components versus one component. In the third model we also evaluate as latent LS factor.

To investigate whether participants in the three countries use the response scales differently, we examine potential disparities in the average vignette ratings across countries, which may suggest variations in scale usage. Importantly, if the SWB ratings in Denmark are due to inconsistent use of numerical scales, one would expect Danes to report higher scores when rating the vignettes compared to Swedes and Australians.

To investigate the potential recall bias and positivity bias, we compare country-specific differences in the mean values of reported mood during the past week with those of reported mood during the current day. We operate under the assumption that if recall biases affect country differences in reported AWB due to recall bias, the differences would be significantly smaller for the latter measure (current day mood). To examine positivity bias, we compare mean values of global LS reports across countries to domain satisfaction reports. If the mean value differences for global reports are larger than domain reports, it may indicate the presence of a positivity bias in the assessment of SWB.

3. Results

3.1 Baseline analysis of mean values of SWB across countries

To investigate the levels of SWB being in the three countries, we begin by analyzing the unadjusted values and comparing levels of SWB using three different measures of AWB and two measures of LS. Figure 1 (below) shows that Denmark has a higher mean value on Bipolar Mood, Positive Affect, SWLS, and Single Item LS, as well as a lower mean value on Negative Affect, compared to both Sweden and Australia. ANOVA tests and pairwise comparisons (Appendix Tables A3 and A4) showed that this advantage for Denmark was statistically significant at the 95% level, while differences between Sweden and Australia were only statistically significant for one of the measures (Bipolar Mood), where Australia scored slightly higher. Results are consistent with previous research indicating that Denmark ranks higher in SWB than Sweden and Australia (Helliwell et al., 2023).

To ensure the robustness of these findings, we also conducted an analysis of covariance (ANCOVA) to calculate mean value differences between the three countries while adjusting for the sociodemographic factors displayed in Appendix Table A1: age, gender, income, and relationship status. The results of the ANCOVA were highly similar to those of the initial analysis, suggesting that the higher levels SWB in Denmark cannot be attributed to these factors (see Appendix Table A5 for a comparison of mean differences with and without covariates).

3.2 Investigating linguistic measurements issues across countries

To investigate linguistic measurement issues in SWB-reports results from our MGCFA-analysis are displayed in Table 2 (below). Within the Multi-Group Confirmatory Factor Analysis (MGCFA) framework, it is typical to evaluate three hierarchical levels of measurement equivalence: configural, metric, and scalar. For our study, it is necessary to establish measurement equivalence across all three levels, as this would provide a foundation for comparing group (country) averages of SWB. Configural equivalence necessitates that the factor loadings' configuration remains consistent across all groups for the items in the model. Metric invariance, on the other hand, is more stringent as it demands equal factor loadings between items and constructs across all groups. This implies that respondents from all groups interpret the latent constructs in the same way. Scalar invariance goes a step further, requiring that the intercepts of the items are also equal across groups, ensuring that the scales have the same zero

point in each group. A model fit decrease in terms of CFI of less than 0.01, when comparing the configural and metric models in large samples (n > 300), typically indicates support for metric invariance, while a similar criterion applies for scalar invariance when comparing metric and scalar models Chen (2007).

Figure 1. *Unadjusted mean values of subjective well-being measures by country*
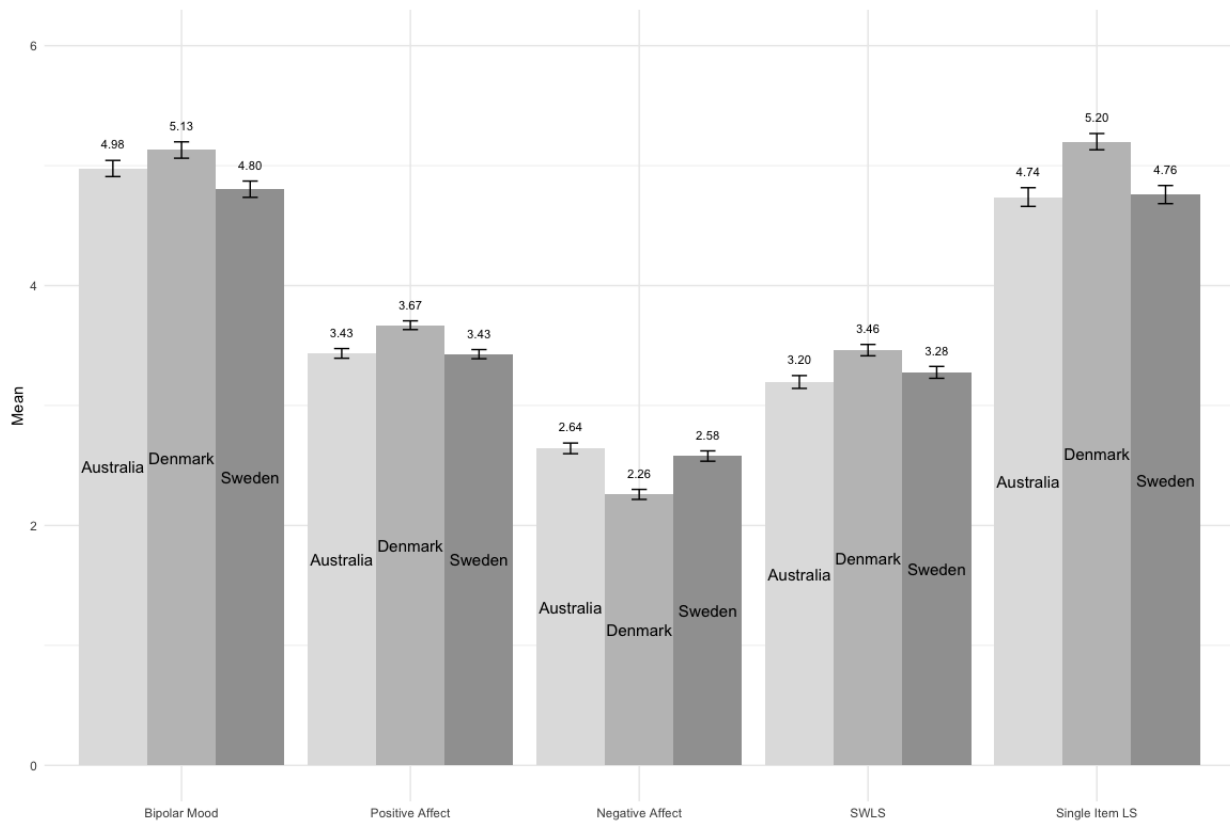
When testing the two-factor model of AWB, both configural and metric invariance were supported across countries. While most factor loadings were relatively similar, the loadings for "Relaxed" and "Engaged" on the PA factor differed more across groups compared to the NA items (see Appendix Table A6). However, the small reduction of CFI from the configural to metric model still met the criteria for metric invariance (Chen, 2007).

In terms of mean values for latent PA and NA, Denmark outperforms both Sweden and Australia, particularly regarding the measure for NA. This means that Denmark has higher latent PA and a lower latent NA, with the differences being most pronounced for NA. However, when testing for scalar invariance, the CFI for this model shows a substantial drop, indicating that the intercepts of the items, when crossing the y-axis for latent PA and NA factors, are not consistent across countries. Consequently, we identified systematic differences in how respondents from the different countries interpret or respond to the items, making it difficult to draw definitive conclusions about Denmark's advantage in PA and NA.

We proceeded to test a single AWB factor, using only the items that capture pure valence through opposing adjectives (sad and happy), combined with the two bipolar mood measures (mood during the past week and mood during today). In this case, the configural, metric, and scalar models all exhibit excellent model fit, with only marginal differences between the various levels of measurement equivalence. As a result, we can confidently compare mean values across countries, given that the items function similarly in all three countries. In line with the results

from measures of PA and NA, Denmark scores significantly higher than Sweden and Australia when AWB is assessed as a bipolar construct. However, Denmark's advantage is not substantial, as the difference is only about 0.1-0.2 standard deviations compared to Sweden and Australia, which, in turn, are not statistically significantly different from each other.

Table 2. *Goodness-of-fit statistics for tests of measurement invariance and differences in latent mean values for Australia, Denmark and Sweden*

	Configural	Metric	Scalar
PA and NA			
CFI	0.945	0.938	0.898
RMSEA	0.108	0.099	0.113
SRMR	0.042	0.059	0.064
Chi square (DF)	376.027 (24)	428.070 (32)	685.380 (40)
Bipolar AWB			
CFI	0.997	0.994	0.986
RMSEA	0.046	0.048	0.061
SRMR	0.010	0.042	0.057
Chi square (DF)	22.010 (6)	46.945 (12)	102.626 (582)
Bipolar AWB and LS			
CFI	0.971	0.969	0.965
RMSEA	0.092	0.086	0.084
SRMR	0.030	0.044	0.052
Chi square (DF)	626.878 (54)	677.048 (66)	770.154 (78)
	Latent mean values		
	Australia	Denmark	Sweden
PA	0.000	0.241***	-0.006
NA	0.000	-0.371***	-0.097
AWB	0.000	0.177***	-0.025
LS	0.000	0.407***	0.091

Note. Comparative fit index (CFI), root mean squared error of approximation (RMSEA), standardized root mean residual (SRMR). Australia serves as the reference country for mean comparisons. *** $p < 0.001$

In the final step of the MGCFA analysis, we compare the bipolar latent measure of AWB with a latent measure of LS. To do this, we expand the model to include the four items designed to capture LS. In this extended model, we achieve measurement equivalence across all three levels, allowing us to confidently compare mean levels across countries for both AWB and LS. The results reveal that Denmark scores higher than Sweden and Australia once again, and that the difference is noticeably larger for LS than for AWB. This observation is supported by the fact that the standardized mean value is almost 0.4 standard deviations higher in Denmark compared to Australia for latent LS, more than twice the mean value difference for AWB.

In summary, our results suggest that linguistic issues do not seem to explain why Denmark exhibits higher levels of SWB than Sweden and Australia, at least not when SWB is measured in terms of bipolar AWB or LS. Hence, the mood questions, affect adjectives, and LS-statements utilized to measure bipolar AWB and LS are interpreted similarly across the three languages.

However, an important nuance in our findings is that the differences between countries are smaller for AWB than for LS. Thus, Danes are on average in a better mood than Swedes and Australians, but only barely so. We remain agnostic regarding differences in PA and NA across countries since we did not obtain measurement invariance for these measures.

3.3 Investigating differences in scale use across countries using vignettes

To assess potential differences in scale usage we analyze emotion focused vignettes. Ratings on the three vignettes are presented in Figure 2. The first vignette (illustrating low levels of AWB) was rated very low in all three countries, with a mean value under 2 in each country. Respondents in the Australian sample rated the vignette significantly more positive (mean = 1.72) than respondents in the Swedish (mean = 1.47) and the Danish sample (mean = 1.32). Contrary to the assumption that the Danes tend to use the positive end of the scale more frequently, the Danish sample was least positive when they judged AWB in the first vignette. ANOVA tests and pairwise comparisons (Appendix Table A7 and A8) showed that these differences in vignette ratings across countries were statistically significant at the 95% level.

Figure 2. *Mean affective well-being scores for anchoring vignettes by country*
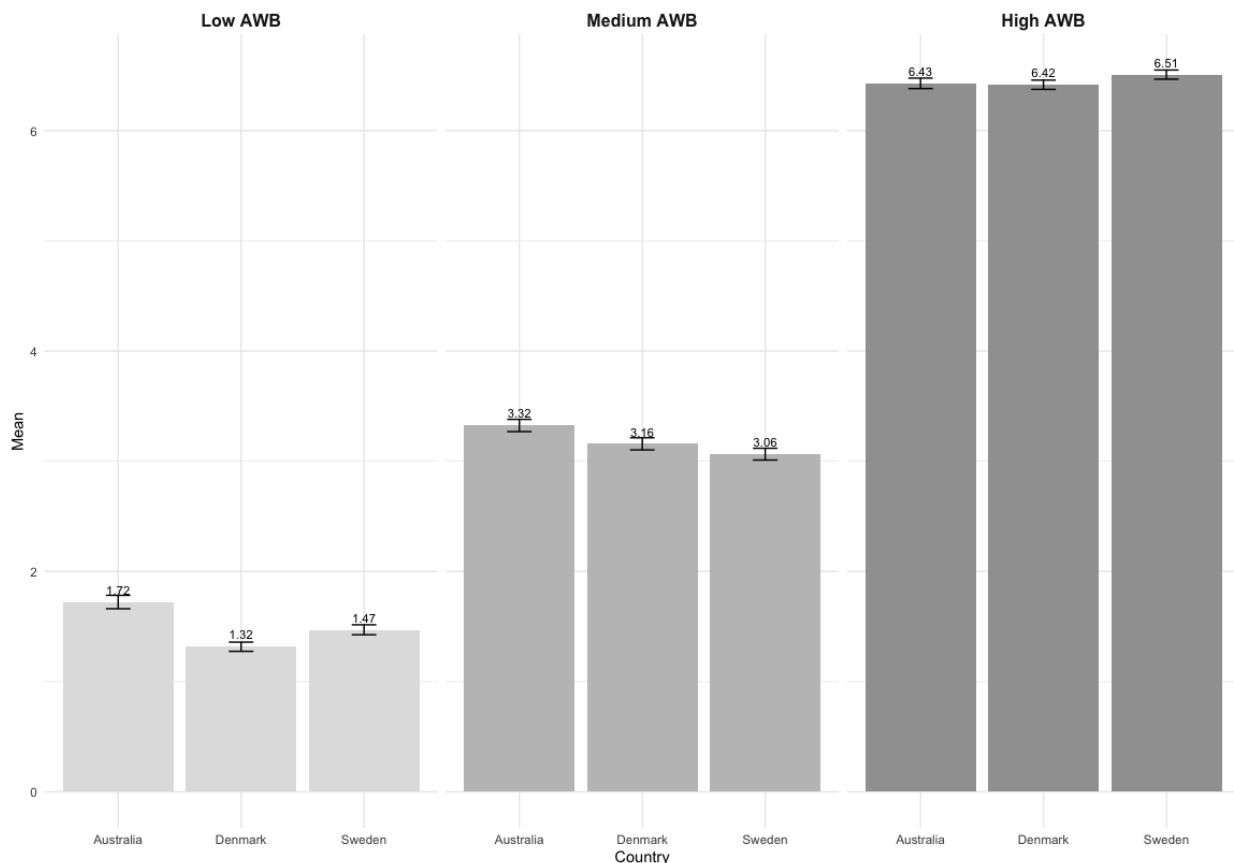
Regarding the second vignette (illustrating levels of AWB in the middle), the mean value was 3.16 in Denmark compared to 3.06 in Sweden and 3.32 in Australia, indicating once again that Australian people, rather than Danish people, tend to give higher (more positive) ratings of the hypothetical person described in the vignette. The Swedes rated this vignette less positively than the Danes, this difference was statistically significant, albeit small (see table A8).

The last vignette, describing a person with high levels of AWB, was rated very high in all three countries with minimal differences in mean values: 6.43 in Australia, 6.42 in Denmark, and 6.51 in Sweden. The higher ratings for Sweden compared to the other two countries were

statistically significant. Thus, even when assessing a person with very high levels of AWB, Danish people were not more prone to use the positive end of the scale compared to people from Australia and Sweden.

In summary, the analysis of vignette ratings across the three countries revealed that Australians tend to provide more positive ratings compared to Swedes and Danes, particularly for the first and second vignettes depicting low and moderate levels of AWB. Contrary to the initial assumption, Danish respondents were not more likely to use the positive end of the scale when assessing AWB. The differences in vignette ratings across countries were generally statistically significant, although the magnitude of these differences was quite small.

3.4 Investigating differences in recall bias across countries

To investigate potential recall bias in reported AWB across three countries – Australia, Denmark, and Sweden – we compared the country-specific differences in the mean values of reported mood during the past week with those of reported mood during the current day, to assess the impact of recall biases on these differences. We operated under the assumption that if recall biases affect country differences in reported AWB, the differences would be significantly smaller for the latter measure (current day mood).

However, our results do not support the assumption that recall biases affect country differences in reported AWB. As illustrated in Figure 3 (below), the mean value differences between Denmark vs Sweden and Australia were not significantly larger for the measure of weekly mood compared to mood today. In fact, the differences were more or less equal across both measures. For example, the difference between Denmark and Sweden for mood during the past week is 5.04 vs. 4.74, while the difference using the mood today measure is 5.22 vs. 4.86. If anything, the reverse is true, as differences are slightly larger for the mood today measure. When we compare Denmark to Australia, the differences in the mood today measure (5.22 vs. 5.08) and the past week measure (5.04 vs. 4.88) is almost the same. These findings suggest that there is no evidence of a recall bias in the reported AWB across these three countries.

Figure 3. *Comparison of mean values for affective well-being (Mood past week & Mood today) by country*
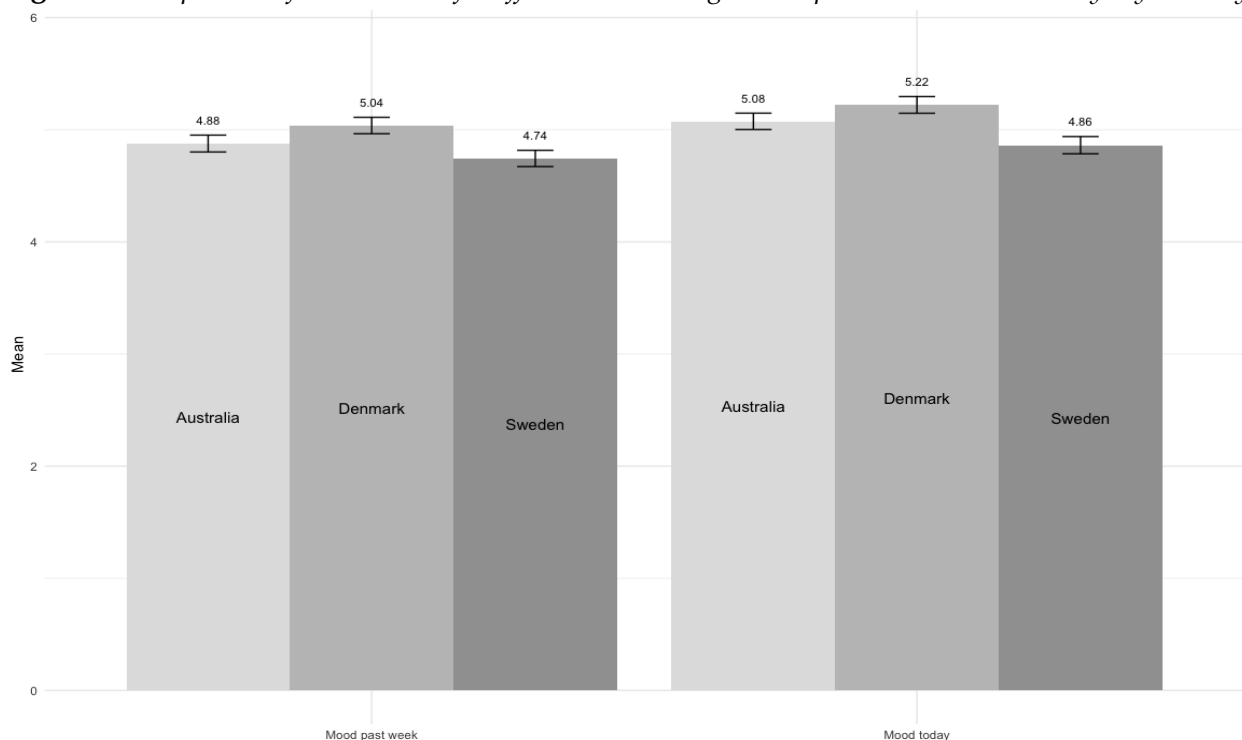
3.5 Investigating differences in positivity bias across countries

To investigate positivity bias we contrast global LS reports with domain-specific satisfaction reports, under the assumption that positivity bias should have a greater impact on global reports than on reports related to specific domains. To compare the mean values of global LS reports across countries with the mean values of domain satisfaction reports, we present z-standardized mean values for each survey item and country in Table 3. The use of z-standardized mean values is necessary because the items assessing domain satisfaction and global LS employed different response scales. The table also includes averaged indices for each type of report (domain satisfaction vs. global LS), enabling a comprehensive comparison of satisfaction levels across countries.

The results reveal that respondents in Denmark consistently scored higher than their Swedish and Australian counterparts across all items. Denmark had a significantly higher average z-score for index measures of both DS and LS compared to Australia (DS: 0. 444 higher, $p < 0.001$; LS: 0.323 higher, $p < 0.001$) and Sweden (DS: 0. 326 higher, $p < 0.001$; LS: 0.256 higher, $p < 0.001$). There were no significant differences in average DS or LS between Sweden and Australia (DS: $p = 0.091$; LS: $p = 0.207$).[4] These findings suggest that individuals in Denmark report higher satisfaction across various life domains and report higher overall LS compared to individuals in Australia and Sweden.

When comparing the differences in DS and LS measures across countries, it is evident that the magnitude of the differences is slightly larger for DS than for LS. The difference in average DS between Denmark and Australia (0.444) is greater than the difference in average LS (0.323). Similarly, the difference in average DS between Denmark and Sweden (0. 326) is larger than the difference in average LS (0.256). This suggests that while Denmark consistently shows higher levels of both DS and LS compared to Australia and Sweden, the gap between countries is slightly more pronounced when considering satisfaction across specific life domains than overall LS, which contradicts the expectation if positivity bias were influencing scores from the Danish respondents.

Table 3. *Comparison of Z-standardized Domain Satisfaction (DS) and Global Life Satisfaction (GLS) Indices across Australia, Denmark, and Sweden*

Construct	Item	Australia	Denmark	Sweden
Domain Satisfaction	Family	-0.213	0.122	0.091
Domain Satisfaction	Finances	-0.164	0.243	-0.079
Domain Satisfaction	Looks	-0.085	0.149	-0.063
Domain Satisfaction	Leisure	-0.095	0.218	-0.123
Domain Satisfaction	Friends	-0.163	0.230	-0.067
Domain Satisfaction	Work/Studies	-0.091	0.147	-0.056
Domain Satisfaction	*DS Index*	*-0.187*	*0.256*	*-0.070*

---

[4] The mean comparisons were performed using Tukey's Honestly Significant Difference test.

Table 3. *Comparison of Z-standardized Domain Satisfaction (DS) and Global Life Satisfaction (GLS) Indices across Australia, Denmark, and Sweden (Cont.)*

Construct	Item	Australia	Denmark	Sweden
Global Life Satisfaction	LS_Nowadays	-0.119	0.222	-0.104
Global Life Satisfaction	LS_Ideal	-0.084	0.151	-0.067
Global Life Satisfaction	LS_Excellent	-0.111	0.125	-0.014
Global Life Satisfaction	LS_Satisfied	-0.147	0.168	-0.021
Global Life Satisfaction	*GLS Index*	*-0.130*	*0.193*	*-0.063*

Note. LS_Nowadays: "How satisfied are you with your life as a whole nowadays?"; LS_Ideal: "In most ways my life is close to my ideal"; LS_Excellent: "The conditions of my life are excellent"; LS_Satisfied: "I am satisfied with my life".

4. General discussion

The primary objective of this study was to explore cross-country variations in SWB while effectively controlling for four key measurement issues: linguistic/translation issues, response scale bias, recall bias and positivity bias. We focused specifically on Denmark, given its reputation as one of the happiest countries in the world, even when compared to Sweden and Australia, two nations that consistently rank lower in SWB assessments despite performing similarly across various cross-national determinants of SWB. Our primary focus was on AWB rather than LS, as we believe this dimension is more important from a normative perspective. Additionally, AWB has largely been neglected in previous studies on cross-country differences in well-being, despite its importance in capturing individuals' quality of life.

Our study showed that Denmark has higher mean values of AWB and LS compared to Australia and Sweden, consistent with prior research. These differences were maintained even after adjustments for age, gender, income, and relationship status. We addressed four measurement issues that might affect these findings. First, potential linguistic issues could affect survey interpretation, but our analysis showed that respondents across all countries understood the survey questions similarly. Thus, Denmark's higher well-being scores are likely not attributable to translation issues. Second, we considered that Danish people might overuse the positive end of the scale. However, after analysing emotion-focused vignettes for a common reference, we found that Danes were not prone to respond more positively. Third, we examined recall bias by comparing past week and current day mood ratings. Our findings indicated that Danes' advantage in AWB was not explained by recall bias. Finally, we investigated positivity bias by analyzing the relationship between global LS judgments and domain-specific satisfaction reports. The results showed that Danes did not overreport their global LS.

Our study also examined the extent to which LS reliably serves as a proxy for AWB in cross-national comparisons, particularly regarding Denmark's high rankings. The results reveal important nuances in this relationship. While Denmark consistently outperformed Sweden and Australia in both AWB and LS measures, the magnitude of these differences varied. Denmark's advantage in latent LS was more than twice that observed in latent AWB. This discrepancy suggests that LS measures may not perfectly proxy AWB. However, it's important to note that the smaller country differences observed for AWB may partly be due to the measure containing

more noise than the LS measure. While we partly accounted for this by using CFA-analysis, there may still be some residual noise in the AWB measure. This difference in variability likely stems from the framing of the AWB items, which focus on the past week and today, potentially introducing more short-term fluctuations. In contrast, the LS items typically reflect judgments about more stable, long-term states or conditions in one's life.

4.1 Limitations

While this study provides valuable insights into the role of measurement issues in explaining Denmark's high levels of SWB, several limitations should be acknowledged. Firstly, the study focused on a limited number of countries - Denmark, Sweden, and Australia. Although these countries were selected based on their performance in various social, economic, and health indicators that are correlated with cross-national differences of SWB, the generalizability of the findings may be limited. Including additional countries with more diverse cultural, social, and economic backgrounds could help to better understand the factors that contribute to cross-country differences in SWB and the role of measurement issues in explaining these differences. Despite this, the focused comparison of the three selected countries is justified. Previous research has raised concerns about the reliability of cross-cultural comparisons of SWB, even among countries that appear similar, underlining the complexity of ensuring comparability across nations.

Secondly, our measure of recall biases was rather rudimentary as we only had one item to measure participants' ratings of their mood during the current day. Future research could benefit from employing more sophisticated methods to measure recall bias in AWB, e.g. by using experience sampling and perhaps contrasting these ratings with a two-week measure of recalled affect. Third, the study relied on self-report measures of AWB. Although we controlled for four different self-report biases, it is still possible that other biases could affect the comparability of well-being measures across countries. Future research should consider exploring alternative methods for assessing AWB, such as behavioural measures or biomarkers like blood pressure or cortisol levels (c.f. Steptoe et al., 2009).

Thirdly, while we have examined linguistic inconsistencies, response scale bias, recall bias, and positivity bias as distinct measurement issues, we acknowledge the interconnections among them. For example, linguistic nuances may influence vignette interpretation, potentially affecting response scale bias assessments. Similarly, positivity bias could impact how individuals use response scales. Cultural differences in emotional expression might simultaneously influence recall bias and LS judgments. Although we addressed each issue individually, we recognize that this approach cannot completely rule out potential overlaps or interactions between these measurement issues. Still, the consistent results reported across all methods strengthen the validity of the findings.

4.2 Future Research

The findings of this study raise several important questions and suggest promising avenues for future research in the field of SWB, particularly in the context of cross-country comparisons. Addressing these areas will not only help to refine our understanding of the factors contributing to differences in SWB across countries but also inform policy interventions aimed at enhancing well-being on a global scale.

Considering that measurement issues may not account for Denmark's high levels of SWB, it is imperative for future research to investigate substantive factors that could be influencing this phenomenon. For instance, an exploration into cultural elements, such as 'hygge' and 'frisind',

which might significantly contribute to the elevated well-being observed in Denmark compared to other countries, would be valuable. However, to our knowledge, there are currently no survey instruments available that effectively measure these specific cultural dimensions on an international scale, presenting a challenge in directly testing such hypotheses.

Given the potential biases and inaccuracies inherent in self-report measures, future research could explore alternative methods for assessing SWB across countries. Ecological momentary assessment or experience sampling methods, for instance, involve repeated assessments of individuals' well-being in real-time and in their natural environment, and could offer a more accurate and nuanced understanding of AWB across different contexts. While there are no existing experience sampling studies that focus specifically on national differences in SWB the use of smartphones for cost-effective data collection could help address this issue and open up new avenues of research. By addressing these areas in future research, we can continue to refine our understanding of the complex factors that contribute to cross-country differences in SWB.

4.3 Conclusion

In conclusion, our study sheds light on the complex nature of cross-country differences in SWB by examining and controlling for four key measurement biases. Our findings reveal that, no such biases could be detected, Danes consistently report higher SWB compared to their Swedish and Australian counterparts. This observation indicates that Denmark's elevated SWB rankings cannot be solely attributed to measurement biases, but rather suggest the presence of genuine differences in well-being. As a result, our research highlights the necessity of exploring other contributing factors, such as social and cultural influences, when investigating cross-country disparities in SWB.

Authors

Filip Fors Connolly
Department of Sociology, Umeå University, Umeå, Sweden
https://orcid.org/0000-0002-3857-4398
filip.fors@umu.se

Johannes Frech
Independent Researcher, Umeå, Sweden

Bengt Brülde
Department of Philosophy, Linguistics, and Theory of Science, University of Gothenburg, Gothenburg, Sweden
https://orcid.org/0000-0002-4481-7253

Christopher Kullenberg
Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Gothenburg, Sweden
https://orcid.org/0000-0002-1577-3570

Conflict of interest statement

The authors report no conflicts of interest.

Author contributions statement

Data availability statement

The data utilized in this study can be accessed on request from the corresponding author.

Acknowledgements

Publishing Timeline

References

Angelini, V., Cavapozzi, D., Corazzini, L., & Paccagnella, O. (2014). Do Danes and Italians rate life satisfaction in the same way? Using vignettes to correct for individual-specific scale biases. *Oxford Bulletin of Economics and Statistics, 76*(5), 643-666. https://doi.org/10.1111/obes.12039

Argyle, M. (2001). *The Psychology of Happiness*, (2nd ed.).. London: Routledge.

Billiet, J. (2003). Cross-cultural equivalence with structural equation modeling. In J. A. Harkness, F. J. R. Van de Vijver, & P. Ph. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 53-79). John Wiley & Sons.

Biswas-Diener, R., Vittersø, J., & Diener, E. (2010). The Danish effect: Beginning to explain high well-being in Denmark. *Social Indicators Research, 97(2)*, 229-246. https://doi.org/10.1007/s11205-009-9499-5

Bille, M. (2015). Lighting up cosy atmospheres in Denmark. *Emotion, Space and Society, 15*, 56-63. https://doi.org/10.1016/j.emospa.2013.12.008

Boas, T. C., Christenson, D. P., & Glick, D. M. (2018). Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. *Political Science Research and Methods, 8(2)*, 232–250. https://doi.org/10.1017/psrm.2018.28

Brülde, B. (2007). Happiness theories of the good life. *Journal of Happiness Studies, 8(1)*, 15–49. https://doi.org/10.1007/s10902-006-9003-8

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.

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the Experience Sampling Method. In *Flow and The Foundations of Positive Psychology* (pp. 35–54). Springer. https://doi.org/10.1007/978-94-017-9088-8_3

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E., Napa-Scollon, C. K., Oishi, S., Dzokoto, V., & Suh, E. M. (2000). Positivity and the construction of life satisfaction judgments: Global happiness is not the sum of its parts. *Journal of Happiness Studies, 1*(2), 159-176. https://doi.org/10.1037/0033-2909.125.2.276

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2009). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143-156. https://doi.org/10.1007/s11205-009-9493-y

Fors, F., & Kulin, J. (2016). Bringing affect back in: Measuring and comparing subjective well-being across countries. *Social Indicators Research, 127*(1), 323-339. https://doi.org/10.1007/s11205-015-0947-0

Glatzer, W. (2000). Happiness: Classic theory in the light of current research. *Journal of Happiness Studies, 1*(4), 501-511. https://doi.org/10.1023/A:1011550727571

Helliwell, J. F., Huang, H., & Wang, S. (2019). Changing world happiness. In J. F. Helliwell, R. Layard, & J. Sachs (Eds.), *World Happiness Report 2019* (pp. 10–45). Sustainable Development Solutions Network.

Helliwell, J. F., Huang, H., Norton, M., Goff, L., & Wang, S. (2023). World Happiness, Trust and Social Connections in Times of Crisis. In *World Happiness Report 2023* (11th ed., Chapter 2). Sustainable Development Solutions Network.

Heen, M. S., Lieberman, J. D., & Miethe, T. D. (2014). A comparison of different online sampling approaches for generating national samples. *Center for Crime and Justice Policy, 1*, 1–8.

Jahan, S., Jespersen, E., Mukherjee, S., & et al. (2017). *Human Development Report 2016: Human Development for Everyone.* United Nations Development Programme.

Kahneman, D., & Riis, J. (2005). Living and thinking about it: Two perspectives on life. In F. A. Huppert, N. Baylis, & B. Keverne (Eds.), *The Science of Well-Being* (pp. 285–304). Oxford University Press.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306*(5702), 1776–1780. https://doi.org/10.1126/science.1103572

Kahneman, D., Schkade, D. A., Fischler, C., Krueger, A. B., & Krilla, A. (2010). The structure of well-being in two cities: Life satisfaction and experienced happiness in Columbus, Ohio; and Rennes, France. In E. Diener, J. Helliwell, & D. Kahneman (Eds.), *International Differences in Well-being* (pp. 16–33). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732739.003.0002

King, G., Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis, 15*(1), 46-66. https://doi.org/10.1093/pan/mpl011

Lolle, H. L., & Andersen, J. G. (2016). Measuring happiness and overall life satisfaction: A Danish survey experiment on the impact of language and translation problems. *Journal of Happiness Studies, 17*(4), 1337-1350. https://doi.org/10.1007/s10902-015-9646-4

OECD. (2011). *How's life? Measuring Well-Being*. OECD Publishing. https://doi.org/10.1787/9789264121164-en

Oishi, S. (2010). Culture and well-being: Conceptual and methodological issues. In E. Diener, D. Kahneman, & J. F. Helliwell (Eds.), *International Differences in Well-Being* (pp. 34–69). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732739.003.0003

R Core Team (2023). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. https://www.R-project.org/

Raudenská, P. (2020). The cross-country and cross-time measurement invariance of positive and negative affect scales: Evidence from European Social Survey. *Social Science Research, 86*, 102369. https://doi.org/10.1016/j.ssresearch.2019.102369

Salomon, J. A., Tandon, A., & Murray, C. J. (2004). Comparability of self-rated health: Cross-sectional multi-country survey using anchoring vignettes. *BMJ, 328*(7434), 258. https://doi.org/10.1136/bmj.37963.691632.44

Saris, W.E. (2001) What influences subjective well-being in Russia? *Journal of Happiness Studies*, Vol. 2: 137-146. https://doi.org/10.1023/A:1011556429502

Saris, W. E., & Gallhofer, I. (2007). Can questions travel successfully? In Jowell, et al. (Eds.), *Measuring Attitudes Cross-Nationally* (pp. 53–79). London: Sage.

Sortheix, F. M., & Weber, W. (2023). Comparability and reliability of the positive and negative affect scales in the European Social Survey. *Frontiers in Psychology, 14*.

Steptoe, A., Dockray, S., & Wardle, J. (2009). Positive affect and psychobiological processes relevant to health. *Journal of Personality, 77*(6), 1747-1776.

Veenhoven, R. (1984) *Conditions of Happiness*. Dordrecht: D. Reidel.

Västfjäll, D., Friman, M., Gärling, T., & Kleiner, M. (2002). The measurement of core affect: A Swedish self-report measure. *Scandinavian Journal of Psychology, 43*(1), 19–31. https://doi.org/10.1111/1467-9450.00265

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science, 14*(5), 520-524.

Appendix

Table A1. *Descriptive statistics for sociodemographic variables in Australia, Denmark and Sweden*

		Australia		Denmark		Sweden	
		n	%	n	%	n	%
Age (years)	18–24	150	11.9	169	13.4	115	9.1
	25–34	216	17.1	174	13.8	231	18.3
	35–44	210	16.7	243	19.3	228	18.1
	45–54	240	19	227	18	239	19
	55–64	276	21.9	233	18.5	216	17.1
	65 or older	168	13.3	214	17	231	18.3
Gender	Male	581	46.1	585	46.4	589	46.7
	Female	679	53.9	675	53.6	671	53.3
Household income (net $USD)	$0 to $24,999	240	19	187	14.8	261	20.7
	$25,000 to $49,999	315	25	285	22.6	390	31.0
	$50,000 to $74,999	288	22.9	331	26.3	326	25.9
	$75,000 to $99,999	179	14.2	214	17	184	14.6
	$100,000 or more	238	18.9	243	19.3	99	7.9
Relationship status	Cohabiting	773	61.3	774	61.4	753	59.8
	Has partner / not cohabiting	79	6.3	105	8.3	96	7.6
	Single (no partner)	408	32.4	381	30.2	411	32.6

Table A2. *Sample and population statistics for gender and age*

	Australia			Denmark			Sweden		
	Sample	Population	Difference	Sample	Population	Difference	Sample	Population	Difference
Gender									
Women	53.9%	50.2%	3.7%	53.6%	50.6%	3%	53.3%	50.6%	2.7%
Men	46.1%	49.8%	3.7%	46.4%	49.4%	3%	46.7%	49.4%	2.7%
Age									
18-24	11.9%	12.2%	-0.3%	13.4%	11.7%	1.7%	9.1%	10.7%	-1.6%
25-34	17.1%	18.9%	-1.7%	13.8%	15.3%	-1.5%	18.3%	17.0%	1.3%
35-44	16.7%	17.3%	-0.7%	19.3%	16.1%	3.2%	18.1%	15.9%	2.2%
45-54	19.0%	16.7%	2.3%	18.0%	17.8%	0.3%	19.0%	16.7%	2.2%
55-64	21.9%	14.7%	7.2%	18.5%	15.3%	3.2%	17.1%	14.5%	2.7%
65+	13.3%	20.2%	-6.9%	17.0%	23.8%	-6.9%	18.3%	25.1%	-6.8%

Table A3. *ANOVA summary for unadjusted mean values of subjective well-being measures across Australia, Denmark, and Sweden*

Construct	Df	Sum Sq	Mean Sq	F value	p-value
Bipolar Mood	2	67.268	33.634	22.367	< 0.001
	3777	5679.574	1.504		
Positive Affect	2	47.644	23.822	49.389	< 0.001
	3777	1821.800	0.482		
Negative Affect	2	106.574	53.287	87.691	< 0.001
	3777	2295.158	0.608		
SWLS	2	46.773	23.387	28.360	< 0.001
	3777	3114.656	0.825		
Single Item LS	2	171.570	85.785	47.960	< 0.001
	3777	6755.825	1.789		

Table A4. *Pairwise comparisons of unadjusted mean values of subjective well-being measures across Australia, Denmark, and Sweden*

Construct	Comparison	Diff.	Lower	Upper	p-adj.
Bipolar Mood	Denmark-Australia	0.154	0.039	0.269	0.005
	Sweden-Australia	-0.173	-0.287	-0.058	0.001
	Sweden-Denmark	-0.327	-0.441	-0.212	<0.001
Positive Affect	Denmark-Australia	0.235	0.170	0.300	<0.001
	Sweden-Australia	-0.006	-0.071	0.059	0.971
	Sweden-Denmark	-0.241	-0.306	-0.176	<0.001
Negative Affect	Denmark-Australia	-0.384	-0.457	-0.311	<0.001
	Sweden-Australia	-0.064	-0.137	0.009	0.098
	Sweden-Denmark	0.320	0.247	0.393	<0.001
SWLS	Denmark-Australia	0.266	0.181	0.350	<0.001
	Sweden-Australia	0.080	-0.005	0.165	0.069
	Sweden-Denmark	-0.185	-0.270	-0.101	<0.001
Single Item LS	Denmark-Australia	0.462	0.337	0.587	<0.001
	Sweden-Australia	0.021	-0.104	0.146	0.921
	Sweden-Denmark	-0.441	-0.566	-0.316	<0.001

Note. The pairwise comparisons presented in the table were derived from a one-way analysis of variance (ANOVA) conducted on the well-being measures across the three countries (Australia, Denmark, and Sweden). The ANOVA was followed by post-hoc tests using Tukey's Honest Significant Difference (HSD) method to determine statistically significant differences between country pairs. The table reports the mean differences (Diff), lower (Lower) and upper (Upper) bounds of the 95% confidence intervals, and adjusted p-values (p-adj) for each pairwise comparison.

Table A5. *Pairwise comparisons of unadjusted mean values of SWB measures across Australia, Denmark, and Sweden with and without covariates*

Construct	Comparison	Diff.	Diff. w/ cov.
Bipolar Mood	Denmark-Australia	0.154	0.113
	Sweden-Australia	-0.173	-0.167
	Sweden-Denmark	-0.327	-0.281
Positive Affect	Denmark-Australia	0.235	0.214
	Sweden-Australia	-0.006	0.000
	Sweden-Denmark	-0.241	-0.214
Negative Affect	Denmark-Australia	-0.384	-0.363
	Sweden-Australia	-0.064	-0.073
	Sweden-Denmark	0.320	0.290
SWLS	Denmark-Australia	0.266	0.221
	Sweden-Australia	0.080	0.111
	Sweden-Denmark	-0.185	-0.110
Single Item LS	Denmark-Australia	0.462	0.405
	Sweden-Australia	0.021	0.042
	Sweden-Denmark	-0.441	-0.363

Note. The pairwise comparisons presented in the table were derived from two separate analyses of the well-being measures across the three countries (Australia, Denmark, and Sweden). The first analysis (Diff.) was a one-way analysis of variance (ANOVA) without covariates, followed by post-hoc tests using Tukey's Honest Significant Difference (HSD) method (see Table A4). The second analysis (Diff. w/cov.) was an analysis of covariance (ANCOVA) that included gender, age, income, and partner status as covariates, followed by pairwise comparisons adjusted for these covariates using the emmeans package in R. The table reports the mean differences for each pairwise comparison from both analyses, allowing for a comparison of the results with and without adjusting for the covariates.

Table A6. *Factor loadings for indicators of latent factors of PA/NA (1), Bipolar Mood (2) and Bipolar Mood and LS (3)*

		Group AUSTRALIA				Group DENMARK				Group SWEDEN			
	Model 1	Esti-mate	S.E.	Est./S.E.	P-Value	Esti-mate	S.E.	Est./S.E.	P-Value	Esti-mate	S.E.	Est./S.E.	P-Value
	HAPPY	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000
PA	RELAXED	0.974	0.043	22.818	<0.001	0.712	0.038	18.873	<0.001	0.849	0.047	17.895	<0.001
	ENGAGED	0.518	0.037	13.930	<0.001	0.663	0.035	19.084	<0.001	0.696	0.037	18.716	<0.001
	SAD	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000
NA	WORRIED	0.965	0.044	21.707	<0.001	0.865	0.043	20.136	<0.001	0.991	0.049	20.096	<0.001
	BORED	0.513	0.041	12.581	<0.001	0.484	0.039	12.495	<0.001	0.599	0.045	13.265	<0.001
	Model 2	Esti-mate	S.E.	Est./S.E.	P-Value	Esti-mate	S.E.	Est./S.E.	P-Value	Esti-mate	S.E.	Est./S.E.	P-Value
	MOOD_WEEK	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000
Bipolar Mood	MOOD_TODAY	0.913	0.032	28.282	<0.001	0.851	0.028	30.347	<0.001	0.870	0.032	26.895	<0.001
	HAPPY	0.613	0.023	27.147	<0.001	0.483	0.018	27.226	<0.001	0.497	0.019	25.782	<0.001
	SAD	-0.590	0.026	-23.064	<0.001	-0.528	0.021	-24.660	<0.001	-0.548	0.024	-22.785	<0.001
	Model 3	Esti-mate	S.E.	Est./S.E.	P-Value	Esti-mate	S.E.	Est./S.E.	P-Value	Esti-mate	S.E.	Est./S.E.	P-Value
	HAPPY	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000
Bipolar Mood	SAD	-0.922	0.037	-24.611	<0.001	-1.057	0.045	-23.604	<0.001	-1.014	0.045	-22.674	<0.001
	MOOD_WEEK	1.428	0.049	28.962	<0.001	1.821	0.061	29.639	<0.001	1.685	0.058	28.836	<0.001
	MOOD_TODAY	1.356	0.048	28.151	<0.001	1.639	0.063	26.143	<0.001	1.542	0.063	24.617	<0.000
	GlobalLS	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000	1.000	0.000	999.000	999.000
LS	SWLS1	0.740	0.021	35.928	<0.001	0.748	0.022	34.260	<0.001	0.676	0.018	37.225	<0.001
	SWLS2	0.679	0.020	34.801	<0.001	0.751	0.021	36.210	<0.001	0.609	0.019	32.135	<0.001
	SWLS3	0.779	0.018	42.707	<0.001	0.768	0.019	40.222	<0.001	0.728	0.017	43.277	<0.001

Table A7. *ANOVA summary table for cross-country comparison of vignettes*

Variable	Df	Sum Sq	Mean Sq	F value	p-value
Low AWB	2	105.621	52.810	64.649	< 0.001
	3777	3085.341	0.817		
Medium AWB	2	43.812	21.906	22.400	< 0.001
	3777	3693.601	0.978		
High AWB	2	6.273	3.137	5.016	0.007
	3777	2361.867	0.625		

Table A8. *Pairwise comparisons of vignettes across Australia, Denmark, and Sweden*

Vignette	diff	lwr	upr	p adj	comparison
Low AWB	-0.406	-0.490	-0.321	<0.001	Denmark-Australia
	-0.252	-0.336	-0.167	<0.001	Sweden-Australia
	0.154	0.070	0.238	<0.001	Sweden-Denmark
Medium AWB	-0.167	-0.259	-0.074	<0.001	Denmark-Australia
	-0.260	-0.353	-0.168	<0.001	Sweden-Australia
	-0.094	-0.186	-0.001	0.046	Sweden-Denmark
High AWB	-0.013	-0.087	0.061	0.914	Denmark-Australia
	0.079	0.005	0.153	0.032	Sweden-Australia
	0.092	0.018	0.166	0.010	Sweden-Denmark

Note. The pairwise comparisons presented in the table were derived from a one-way analysis of variance (ANOVA) conducted on the well-being measures across the three countries (Australia, Denmark, and Sweden). The ANOVA was followed by post-hoc tests using Tukey's Honest Significant Difference (HSD) method to determine statistically significant differences between country pairs. The table reports the mean differences (Diff), lower (Lower) and upper (Upper) bounds of the 95% confidence intervals, and adjusted p-values (p-adj) for each pairwise comparison.

Table A9. *ANOVA summary table for cross-country comparison of mood measures across time frames*

Variable	Df	Sum Sq	Mean Sq	F value	p-value
Mood Past Week	2	54.497	27.248	15.343	< 0.001
	3777	6707.815	1.776		
Mood Today	2	82.343	41.172	22.476	<0 .001
	3777	6918.862	1.832		

Table A10. *Pairwise comparisons of mood measures across Australia, Denmark, and Sweden*

Variable	diff	lwr	upr	p adj	comparison
Mood Past Week	0.161	0.037	0.286	0.007	Denmark-Australia
	-0.133	-0.257	-0.008	0.034	Sweden-Australia
	-0.294	-0.418	-0.169	0.000	Sweden-Denmark
Mood Today	0.147	0.020	0.273	0.018	Denmark-Australia
	-0.213	-0.339	-0.086	0.000	Sweden-Australia
	-0.360	-0.486	-0.233	0.000	Sweden-Denmark

Note. The pairwise comparisons presented in the table were derived from a one-way analysis of variance (ANOVA) conducted on the well-being measures across the three countries (Australia, Denmark, and Sweden). The ANOVA was followed by post-hoc tests using Tukey's Honest Significant Difference (HSD) method to determine statistically significant differences between country pairs. The table reports the mean differences (Diff), lower (Lower) and upper (Upper) bounds of the 95% confidence intervals, and adjusted p-values (p-adj) for each pairwise comparison.