CrossMark

The Impact of Scale Transformations on National Subjective Well-Being Scores

Cassondra Batz[1] · Scott Parrigon[1] · Louis Tay[1]

© Springer Science+Business Media Dordrecht 2015

Abstract Nation-level subjective well-being (SWB) indicators (e.g., happiness and life satisfaction) have received increased interest among researchers and policy makers. This increased interest has led to questions about the measurement and comparability of SWB scores. In this paper, we examined the comparability of scores derived from scale transformations (e.g., the "linear stretch method"): a procedure that places SWB scales with different scale points (e.g., 3-points, 4-points) on a common metric. Controlling for nation-level and temporal trends, our results reveal that scale transformations potentially lead to biases for both happiness and life satisfaction scales. Moreover, these potential biases may affect substantive research. This is illustrated when we demonstrate that scale transformations may substantially attenuate the relation between national Gross Domestic Product per capita and happiness. Implications of these findings and limitations of the current study are discussed.

Keywords Subjective well-being · Measurement · Scale transformations · Measurement equivalence

1 Introduction

While societies have traditionally relied on economic indicators to measure and track societal progress, there is now increased recognition that indicators beyond these can provide important and relevant information for policy making; one such indicator being subjective well-being (SWB) (Diener 2000; Diener and Seligman 2004). Research has shown that SWB is a reliable and valid indicator (Diener et al. 2013) and is aligned with

✉ Cassondra Batz
  cbatz@purdue.edu

[1] Department of Industrial-Organizational Psychology, Purdue University, 703 Third Street, West Lafayette, IN 47907, USA

important (and valued) societal characteristics such as higher national wealth (Diener et al. 2013), low national corruption (Tay et al. 2014), and higher levels of democracy and freedom (Inglehart et al. 2008).

As more nations are adopting measures of SWB (e.g., National Research Council 2013; OECD 2013), it has been questioned whether national SWB scores are comparable across nations and across time for various purposes such as benchmarking and comparing SWB standards (e.g., Judge and Kammeyer-Mueller 2011; Vennhoven 2014). Generally, two distinct approaches—but not mutually exclusive—can be used for examining differences in national SWB. One approach would be to use uniform scales across nations to ensure that scores are at least comparable *prima facie* as there may be other issues such as the measurement equivalence of SWB scores that need to be further addressed. However, the use of standardized SWB scales across nations has not been widely implemented. Instead, different SWB scales are frequently used across nations (e.g., different wording or scaling). In order to utilize this available data despite these differences in the measures, a common approach involves homogenizing responses to survey questions that differ in some capacity, but involve the same matter of interest (de Jonge et al. 2014). This approach attempts to create equivalence between SWB scales by statistically modifying, or transforming, the scales. An example of this approach is transforming a score from a 4-point scale to a 10-point scale for purposes such as tracking national happiness over time (e.g., Easterlin 1974, 1995) or for comparing SWB across nations (Veenhoven 2014).

However, a critical question arises from the homogenizing approach: do scale transformations produce equivalent scores? If scale transformations do not create equivalence, it could inadvertently impact the comparison of national SWB; further, it may impinge on substantive research and lead to less accurate findings. Increasingly, due growing interest in SWB, researchers use SWB data sampled from multiple nations with different SWB scales. This typically requires scale transformations for comparison purposes (see Table 1; e.g., Arampatzi et al. 2015; Fischer and Boer 2011; Fischer and Van de Vliert 2011; Frey and Stutzer 2002; Graham 2005; Ovaska and Takashima 2010; Steel and Ones 2002). Practically, if scale transformations produce non-equivalent scores, then extra care would be required when making inferences from transformed scores. However, if scale transformations produce equivalent scores, then this practice would be defensible for research and policy purposes.

The current study seeks to answer the aforementioned question by investigating whether scale transformations lead to score comparability in national SWB scores, including both happiness and life satisfaction measures. Specifically, we examine score comparability by assessing whether transformations from different scale types (e.g., 3-point scale, 4-point scale, etc.) lead to substantial score differences while controlling for nation-level and temporal effects.

Our paper is structured as such: (1) we begin by reviewing the current literature on general scaling and measurement issues in SWB research; (2) we then review the specific issue of scale transformations on national accounts of SWB; followed by an (3) empirical examination of the effect scale transformations via the "linear stretch method" have on an archival dataset of national SWB scores. (4) Based on these results, we examine the impact of these potential biases on substantive research by examining the relation between national Gross Domestic Product (GDP) and SWB. (5) Finally, we broadly discuss the implications of our findings for the measurement and tracking of societal well-being.

Table 1 Sampling of cross-national SWB studies

Article	Question posed	Findings
Graham (2005)	What are the discrepancies between well-being measures and income measures?	Reported well-being in Latin America and Russia have discrepancies between well-being and income- or expenditure-based measures, taking this into account may improve understanding of development outcomes
Steel and Ones (2002)	Do national personality traits explain national subjective well-being?	Neuroticism and extraversion were correlated with national SWB and incrementally predicted SWB above gross national product per capita
Fischer and Van de Vliert (2011)	How does culture, climate and wealth interact and influence psychological processes? (i.e. Ill-being)	Threatening climatic demands not met by resources result in ill-being such as health complaints, burnout, anxiety, and depression. The interaction account for 35 % of the variation in overall subjective ill-being
Frey and Stutzer (2002)	What if anything can be gained from using SWB in national policy and economic applications?	SWB is a satisfactory empirical approximation to individual utility and can contribute to policy decisions. Economic variables affect happiness and institutional factors
Fischer and Boer (2011)	What is more important: to provide citizens with more money or with more autonomy to improve SWB?	Individualism is a better predictor of SWB, than wealth. Wealth may influence well-being only via its effect on individualism
Arampatzi et al. (2015)	Are employed individuals affected by economic crisis?	Poor economic conditions in a country are negatively associated with life satisfaction of employees dependent upon their personal financial situation and outlook
Ovaska and Takashima (2010)	What factors impact the inequality of happiness in a nation?	Factors such as individual income, quality of health, and levels of institutional qualities tend to be most important in explaining inequalities in well-being

1.1 General Scaling Issues in SWB Research

Recently there have been a number of studies that have focused their attention on general scaling issues involving SWB measures. For example, research has explored the comparability of balanced versus unbalanced designs for happiness response scales (Liao 2013), the comparability and impact of the varying terminology in happiness scales (Kalmijn et al. 2010), the comparability between translated versions of scales from one language to another (Extremera and Fernandez-Berrocal 2013), and the comparability of single item scales and multiple item scales (Abdel-Khalek 2006). Generally, these studies have found that subtle changes can dramatically impact scores, but in some instance do not produce an effect. This suggests that the effects of scaling changes may not unilaterally create non-equivalence, and therefore requires further investigation.

1.2 Scale Transformations for Comparisons of National Well-being

In our paper, we focus on the specific issue of scale transformations where national SWB scores are transformed from one Likert-scale point to another for the purposes of comparison, which is procedure referred to as the "linear stretch method" (de Jonge et al. 2014). For instance, this occurs when one compares scores derived from a 4-point rating scale to a 10-point rating scale by directly converting the scores from one metric to another. Concretely, a score of 3 on a 1-to-4-point rating scale would be converted to a 7.0 on a 1-to-10-point rating scale. We examine this method because it is commonly used due to ease of implementation. Researchers do not require the use of individual-level data to make these transformations. This method is also incorporated in the World Database of Happiness, a leading data repository for the SWB of nations (Veenhoven 2004, 2014).

As applied to the tracking of national SWB, scale transformations have become prevalent in an effort to make comparisons despite the use of SWB measures with different rating scales. The literature appears to have implicitly accepted that scale transformations provide accurate comparisons (Hagerty and Veenhoven 2003; Veenhoven 1995). However, some researchers have recognized that scale transformations could potentially lead to spurious differences (e.g., Easterlin 2005; de Jonge et al. 2014). Despite this, little research has examined the extent to which scale transformations introduce bias.

In fact, to our knowledge, only two studies have examined the effects of scale transformations on SWB scales, both of which have found evidence that scale transformations produce non-equivalent scores. The first study conducted by Lim (2008) found that scale transformations of individual happiness ratings on 4-point, 5-point, 7-point, and 11-point rating scales were not comparable. Specifically, it was found that the 11-point rating happiness scores were significantly higher than the 4-point and 7-point rating happiness scores following a scale transformation.

More recently, a second study by de Jonge, Veenhoven, and Arends (2014) assessed the effect of transformations on happiness scales using data from the Netherlands between the years 1989 and 2009. They found that the common, more simplistic, methods do not produce comparable results (e.g., the linear stretch method), and that data derived from transformed scales were not comparable. Thus, suggesting that two underlying assumptions of this practice may not be tenable: (1) that there are equal distances among the response options and (2) that the labeling of the response options are unimportant (de Jonge et al. 2014).

To our knowledge, de Jonge et al. (2014) are the first to examine the effects of scale transformations at the national level. This seminal work is important although there are additional key issues that need to be examined. First, their study only looked at the impact of scale transformations on happiness measures, and did not explore the impact of scale transformations on life satisfaction measures, which is another common, key indicator of SWB (Diener et al. 2013). Additionally, the dataset that de Jonge et al. (2014) examined utilized only a single country over a limited amount of time. One difficulty in using any single country to examine the impact of scale transformations is that it is difficult to determine whether the lack of comparability is a result of true changes to SWB in specific years that may correspond to changes in scale response type. Therefore it is critical to extend this to multiple countries over multiple years to determine if the lack of comparability continues to hold when controlling for both nation and temporal effects (i.e., differences in national SWB and changes in SWB over time). In our present study we address these issues and extend past research by (1) exploring the comparability of both happiness and life satisfaction scales post transformation while (2) accounting for nation-

level and temporal effects, which allows us to make a stronger inference about whether scale transformations result in equivalent scores.

1.3 The Present Study

The current study builds on the work by de Jong et al. (2014) who examined the effect of scale transformations on national SWB scores of happiness. To expand on their work, we include data from an extended time frame (1945–2013) and from a larger sample of countries (Happiness $N = 57$; Life Satisfaction $N = 66$) using data from the National Database of Happiness (Veenhoven 2014). Further, we examine two components of subjective well-being: happiness and life satisfaction. In addition, we seek to estimate the scale transformation effects controlling for nation-level and temporal effects to ensure that our estimates of scale transformation effects account for the potentially confounding effects of nations and time.

2 Methods

2.1 Data Source

The World Database of Happiness (Veenhoven 2014) was used for this study. The database was created to serve as an ongoing register of scientific research on the subjective enjoyment of life. It brings together findings that are scattered throughout many studies to provide a basis for synthetic work on the topic of national well-being and address the limitations of current research on well-being by providing a complete inventory of contemporary publications on the subject and three homogenous collections of research findings; two collections of distributional findings (one of which we used for the purpose of our study) and one collection of correlational findings (Veenhoven 2014). (For a comprehensive description of the database see Veenhoven 2004).

Specifically, this database is composed of SWB data compiled between 1945 and 2013 from 164 nations. Generally, the data was collected through cross-national surveys, or by periodic Quality-of-Life surveys within particular nations. This database was compiled by searching through various sources of data such as archived abstracts, library catalogues and data banks, and by directly contacting investigators in the field resulting in a collection of distributional findings of some 3000 general population surveys in nation's studies (Veenhoven 2009).

From the World Database of Happiness (Veenhoven 2014), we extracted data from countries that had longitudinal reports (i.e., more than three time points) for measures of happiness and life satisfaction (see next section for additional information about the measures used). Longitudinal reports are deemed as important to be able to model for national temporal trends over time to more reliably distinguish among years of measurement, the national average over time, and scale type. Based on this criterion, a total of 67 nations that had data collected between the years of 1945 and 2013 was used for our analysis.

2.2 Measures

A variety of SWB measures comprised of four main parts ('overall happiness' measures, 'hedonic level' measures, 'contentment' measures, and 'mixed indicators') was used in the database. Within each of these four categories, the collection of measures was further differentiated based on the question's wording and rating scale length (Veenhoven 2014). For this study we focused on the largest cluster of measures that were categorized as

'Overall Happiness scales,' which contained measures of both happiness and life satisfaction. Though this cluster of scales was not identical in wording and structure, the scales were defined as being equivalent according to Veenhoven's classification. Specifically, scales that tapped into happiness, but used different rating scale points (e.g., 3-point verbal happiness, 4-point verbal happiness, 5-point verbal happiness, 7-point verbal happiness, 10-point verbal happiness, and 11-point verbal happiness) were considered 'happiness' measures (see Table 2 for an example of the wording and the rating scale of each measure). Similarly, scales that tapped into life satisfaction, but used different rating scale points (e.g., 3-point verbal life satisfaction, 4-point verbal life satisfaction, 5-point verbal life satisfaction, 7-point verbal life satisfaction, 10-point verbal life satisfaction, and 11-point verbal life satisfaction) were considered 'life satisfaction' measures (see Table 3 for an example of the wording and the rating scale of each measure).

2.3 Scale Transformations

The scores from the original scales were transformed to a 10-point rating scale, with the exception of the 10-point rating scale, which was used as the baseline. Transformations for the scales with 3, 4, 5, and 7-points were computed using the following equation:

Table 2 Happiness scale label, wording, and options

Scale label	Example wording	Scale options
3-Point verbal Happiness Scale	Taken all together, how would you say things are these days? Would you say that you are…?	1-Not too happy 2-Pretty happy 3-Very happy
4-Point verbal Happiness Scale	Taking all things together, would you say you are…?	1-Not at all happy 2-Not very happy 3-Quite happy 4-Very happy
5-Point verbal Happiness Scale	How happy do you feel as you live now? Please choose one item on this card that comes closest to your feeling	1-Very unhappy 2-Fairly unhappy 3-Neither happy nor unhappy 4-Fairly happy 5-Very happy
7-Point verbal Happiness Scale	If you were to consider your life in general, how happy or unhappy would you say you are, on the whole?	1-Completely unhappy 2-Very unhappy 3-Fairly unhappy 4-Neither happy nor unhappy 5-Fairly happy 6-Very happy 7-Completely happy
10-Point numeral Happiness Scale	Taking all things together on a scale of one to 10, how happy would you say you are?	1-Very unhappy 2…9 10-Very happy
11-Point numeral Happiness Scale	Taking all things together, how happy would you say you are?	0-Extremely unhappy 1…9 10-Extrememly happy

Table 3 Life Satisfaction Scale label, wording, and options

Scale label	Example wording	Scale options
3-Point verbal Life Satisfaction Scale	How satisfied are you with the way you are getting on now…?	3-Very satisfied
		2-All right
		1-Dissatisfied
4-Point verbal Life Satisfaction Scale	How satisfied are you with the life you lead?	4-Very satisfied
		3-Fairly satisfied
		2-Not very satisfied
		1-Not at all satisfied
5-Point verbal Life Satisfaction Scale	How happy do you feel as you live now? Please choose one item on this card that comes closest to your feeling	1-Very unhappy
		2-Fairly unhappy
		3-Neither happy nor unhappy
		4-Fairly happy
		5-Very happy
7-Point verbal Life Satisfaction Scale	Asking all together, how satisfied are you with your life-as-a-whole these days?	7-Completely satisfied
		6-Very satisfied
		5-Satisfied
		4-Satisfied nor unsatisfied
		3-Unsatisfied
		2-Very unsatisfied
		1-Completely dissatisfied
10-Point numeral Life Satisfaction Scale	All things considered, how satisfied are you with your life as-a-whole now?	10-Satisfied
		9…2
		1-Dissatisfied
11-Point numeral Life Satisfaction Scale	All things considered, how satisfied or dissatisfied are you with your life as-a-whole these days?	10-Very satisfied
		9…1
		0-Not satisfied

$$x_{\text{transformed}} = 1 + (x_{\text{original}} - 1) \times [9/(\max_{\text{original}} - 1)],$$

where x_{original} corresponds to the original score, $\max_{\text{original}}$ corresponds to the maximum score on the original scale, and $x_{\text{transformed}}$ corresponds to transformed score on the 10-point rating scale.

Transformations for the 11-point scales (i.e., 0-to-10 point scale) were computed using the following equation:

$$x_{\text{transformed}} = 1 + (x_{\text{original.11}}) \times (9/10),$$

where $x_{\text{original.11}}$ corresponds to the original score on the 11-point scale, and $x_{\text{transformed}}$ corresponds to transformed score on the 10-point rating scale.

2.4 Data Analysis

In order to test the equivalence of transformed scores, we used linear mixed-effects model implemented using the *lme4* package (Pinheiro et al. 2015) in the R statistical environment

(R Core Team 2015). Linear mixed-effects models were chosen as they allow for the modeling of crossed nested effects, such as those that exist within the current data structure. As both the country being sampled, and the year the data are sampled in can potentially confound the results, modeling both of these systematic components within a nested data structure allows these sources of variance to be controlled for.

The need to model this nested data structure was evidenced by improved model fit once the nested structure was accounted for. Bayesian Information Criteria (BIC) values showed large improvements when the nested effects structure was modeled versus when it was not for both happiness and life satisfaction analyses (Δ BIC = 524 and 1947, respectively). All analyses met the assumptions of multilevel modeling (independent and identically normally distributed within-group errors, normally distributed random effects, etc.). Additionally, in order to decrease the influence of outliers and ensure the accuracy of the parameter estimates (Pinheiro and Bates 2000), cases with residuals greater than three standard deviations from the mean were excluded from analysis (9 and 23 cases or 1.21 and 1.39 % of cases, respectively).

We tested the effect of scale transformation through the use of an effects coded scale variable with the 10-point scale set as the reference group (coded as -1), which allows for us to examine if any of the scale means deviate from the grand mean of the scale scores (i.e., the mean across all scale types after controlling for the cross-nested data structure). We would expect that, if simple transformations across scales are equivalent to one another, then there would be little to no deviation from the grand mean of the outcomes (happiness and life satisfaction) that is attributable to scale type (i.e., the scale types would not be significantly different from the grand mean) once the nested effects of both country and year sampled are modeled. However, if the number of points on the scale systematically biases the outcomes to be inflated or deflated compared to that of the other scales, then one or multiple scales would show deviation (as indicated by significant coefficients for that scale type) from the average across all of the scale means (the grand mean of the scales).

3 Results

3.1 Happiness

Figure 1 shows that a potential positive linear or curvilinear relationship may exist between the number of points in a scale and the level of national happiness assesses, with the 11-point happiness scale showing the highest levels of happiness. However, while this initial graphical depiction suggests the possibility of systematic biases in the level of happiness measured by each scale type, these differences could simply be artifacts of either the countries that used each scale type or the years in which these scales were sampled. In other words, these effects might simply be due to the fact that certain countries (the Netherlands, for example) used the 10-point scale or that the 10-point scale happened to be used more frequently during a certain era (a general economic downturn, for example), thus potentially systematically biasing these visualizations. Therefore, in order to draw valid inferences about any potential differences that may exist, we must further explore these relationships within the context of a linear mixed-effects modeling framework that can model (and therefore control for) the effects of both country and the year sampled.
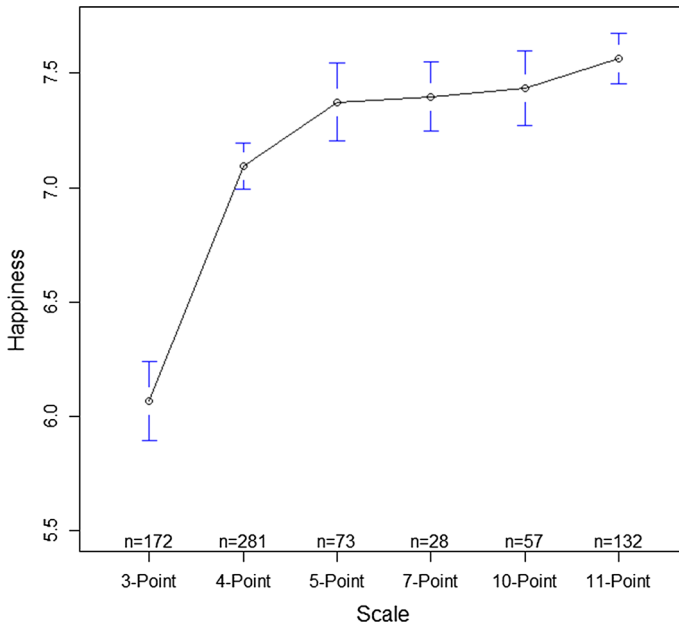
Fig. 1 Happiness scale transformation visual comparison. *Note* The *bars* represent the 95 % confidence intervals for the means. *n* represents the number of times each scale was used

Table 4 Unstandardized linear mixed-effects model results of scale transformation analyses

Fixed effects	Happiness		Life satisfaction	
	b	95 % CI	b	95 % CI
Intercept	6.96[a]	[6.77, 7.16]	6.57[a]	[6.33, 6.82]
3-point scale	−1.37[a]	[− 1.48, −1.27]	−.84[a]	[− 1.06, −.62]
4-point scale	.14[a]	[.05, .22]	.11[a]	[.04, .19]
5-point scale	−.16[a]	[− .28, −.04]	−.05	[− .15, .05]
7-point scale	.40[a]	[.25, .56]	.07	[− .13, .28]
11-point scale	.55[a]	[.45, .64]	.43[a]	[.34, .51]

[a] Confidence interval excludes zero. 95 % confidence intervals represent the bootstrapped estimates of the effects. For Happiness, the number of groups = 57 and the number of observations = 734. For Life Satisfaction, the number of groups = 66 and the number of observations = 1625. Scale was entered as effect coded variables with the 10-point Happiness scale set as the reference group (set at −1). The cross-nested structure of the data within both country and year are modeled

Linear mixed-effect model results, presented in Table 4, reveal that, when accounting for the cross-nesting of happiness measurements within both the country sampled and the year sampled, all of the transformed happiness scales predict systematic deviation attributable to the scale type being used (according to bootstrapped estimates of the effects for each scale type). This suggests that scale transformations may not result in directly comparable scores. Specifically, the 3-point and 5-point scales were found to be significantly lower than the happiness grand mean of 6.96 (1.37 and .16 scale points lower, respectively), with the other scale points (4, 7, and 11) being significantly higher than the

grand mean (.14, .40, and .55 scale points higher, respectively). Overall, these analyses indicate that the use of differing scale points may systematically bias the resultant happiness means and that the standard scale transformations of happiness scales may not result in directly comparable scores.

3.2 Life Satisfaction

A visual inspection of Fig. 2 suggests that a potential positive linear relationship may exist between the number of points in a scale and the resultant level of national life satisfaction. However, as discussed previously, in order to make any valid inferences about potential differences between scales, we need to ensure that these differences are not artifacts of the countries that used each scale or the year they were being sampled. Thus, we explore these relationships within a linear mixed-effects framework that can control for the nesting of each measurement within both country and year.

The results of the linear mixed-effects analysis, seen in Table 4, indicated that the number of scale points used may lead to systematic bias in the ratings of life satisfaction. Specifically, similar to the happiness analyses, the 3-point scale was found to be significantly lower than the grand mean of 6.57 (.84 scale points lower), while the 4 and 7-point scales were higher than the grand mean (.11 and .43 scale points higher, respectively). The 5 and 7-point scales showed a similar pattern of effects as were found in the happiness analysis, these two effects were not significant. Overall, the general pattern of systematic bias found with the happiness scale analyses was generally replicated with life satisfaction. These analyses indicate that transformations of life satisfaction scales may not result in directly comparable scores and that the nature of the biases may hold across type of
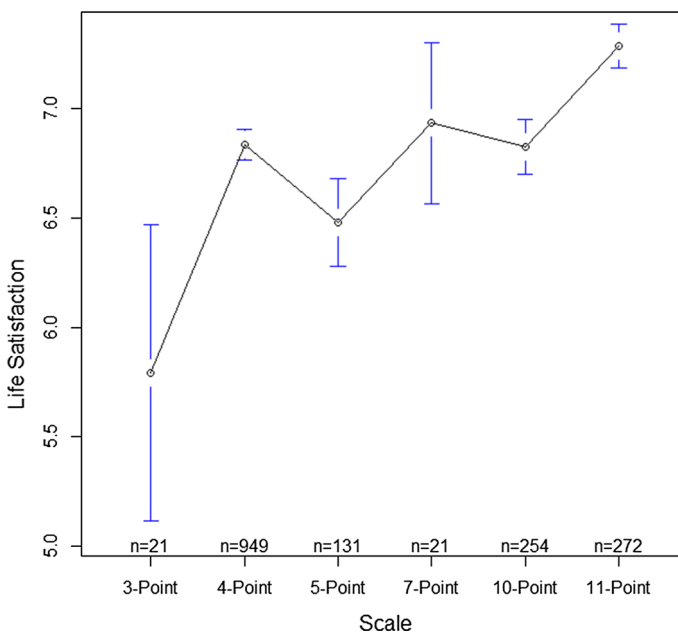

Fig. 2 Life satisfaction scale transformations visual comparison. *Note* The *bars* represent the 95 % confidence intervals for the means. *n* represents the number of times each scale was used

subjective well-being measure (e.g., the 3-point scale shows similar large negative biases for both happiness and life satisfaction).

4 Testing the Substantive Impact of Scale Transformations

Our analyses revealed that scale transformations may not result in comparable scores. However, it is difficult to ascertain the practical impact of these biases from the magnitude of the scale point difference seen in the coefficients in Table 4. As such, we further investigated the potential impact of scale transformations on substantive research such as the relationship between income and SWB, which is of interest to policy makers and researchers alike (Diener and Biswas-Diener 2002; Veenhoven and Vergunst 2014). Specifically, we examine the relationship between national Gross Domestic Product per capita (GDP) and SWB using scale transformations and scale transformations that are corrected for bias (using coefficients derived from our earlier analyses). Assuming that the bias corrections allow for more accurate reflections of national SWB, this analysis can help shed light upon the potential impacts that scale transformations have on substantive findings.

4.1 Data Analysis

GDP per capita was obtained from the World Bank database (2014), which had GDP data for years 1980 through 2014. We use the same SWB data used previously from the World Database of Happiness (Veenhoven 2014) in order to explore these relationships. To assess the potential impact of scale transformations on substantive findings, we compare the results of models where the scale biases are modeled in the analysis (as was done in the previous analyses), versus models where these scale biases are ignored (i.e., not modeled). For ease of interpretability, the standardized coefficients are presented. For happiness, the analyses indicated that allowing the linear slopes for GDP to randomly vary across countries significantly improved model fit for both the corrected and uncorrected models (BICΔ = 17 and 18, respectively). For life satisfaction, allowing both the linear and the quadratic slopes to vary also improved model fit for both the corrected and uncorrected models (BICΔ = 49 and 64, respectively). As such, these were modeled in the analyses.

4.2 Results

The results, provided in Table 5, reveal several important differences emerge when the scale biases are modeled versus when they are not. For happiness, while both the corrected and the uncorrected relationships show positive linear relationships between GDP and happiness, the corrected effects are roughly half the size of the uncorrected effects (β = .33 vs. β = .65). This suggests that, while GDP increases do correspond with increases in happiness, these effects may be approximately half as strong as the raw relationship would indicate. For life satisfaction, the results indicate that GDP is positively related to life satisfaction at roughly the same strength as the corrected, but not the uncorrected, happiness analyses. This suggests that GDP may predict happiness and life satisfaction similarly, as opposed to the roughly doubled strength of happiness that was indicated by the uncorrected analyses. Additionally, the corrected scale analyses suggest a more complex relationship may exist between GDP and life satisfaction such that GDP seems to have diminishing effects at higher levels, as indicated by the significant negative quadratic relationship found in the analyses. A plot of this trend can

Table 5 Standardized linear mixed-effects model results examining the effect of correcting for scale artifacts

Fixed effects	Happiness			
	Uncorrected		Corrected	
	β	95 % CI	β	95 % CI
Step 1				
3-Point scale	–	–	−1.40[a]	[−1.50, −1.30]
4-Point scale	–	–	.14[a]	[.06, .22]
5-Point scale	–	–	−.16[a]	[−.28, −.04]
7-Point scale	–	–	.41[a]	[.24, .58]
11-Point scale	–	–	.56[a]	[.46, .65]
Step 2				
GDP[a]	.65[a]	[.49, .81]	.33[a]	[.22, .44]
Step 3				
GDP quadratic	.02	[−.05, .10]	.02	[−.03, .08]
Fixed effects	Life satisfaction			
	Uncorrected		Corrected	
	c	95 % CI	β	95 % CI
Step 1				
3-Point scale	–	–	−.80[a]	[−1.04, −.55]
4-Point scale	–	–	.11[a]	[.03, .18]
5-Point scale	–	–	−.05	[−.15, .05]
7-Point scale	–	–	.06	[−.12, .25]
11-Point scale	–	–	.41[a]	[.32, .49]
Step 2				
GDP[a]	.30[a]	[.19, .42]	.29[a]	[.18, .41]
Step 3				
GDP quadratic	−.06	[−.08, .02]	−.07[a]	[−.13, −.02]

[a] Confidence interval excludes zero. 95 % confidence intervals represent the bootstrapped estimates of the effects. For Happiness, the number of groups = 57 and the number of observations = 734. For Life Satisfaction, the number of groups = 66 and the number of observations = 1506. For the corrected values, scale type was entered in at Step 1 as effect coded variables with the 10-point Happiness scale set as the reference group (set at −1). The cross-nested structure of the data within both country and year are both modeled

be found in Fig. 3. Together, these results indicate that correcting the scale transformations can reveal important and nuanced relationships that may be hidden by simply using standard scale transformations.

5 Discussion

The present research suggests that scale transformations may produce biases, which aligns with the findings of de Jonge et al. (2014), but also goes beyond their findings. Generally, transformations inflate the means particularly when transforming shorter scales into longer
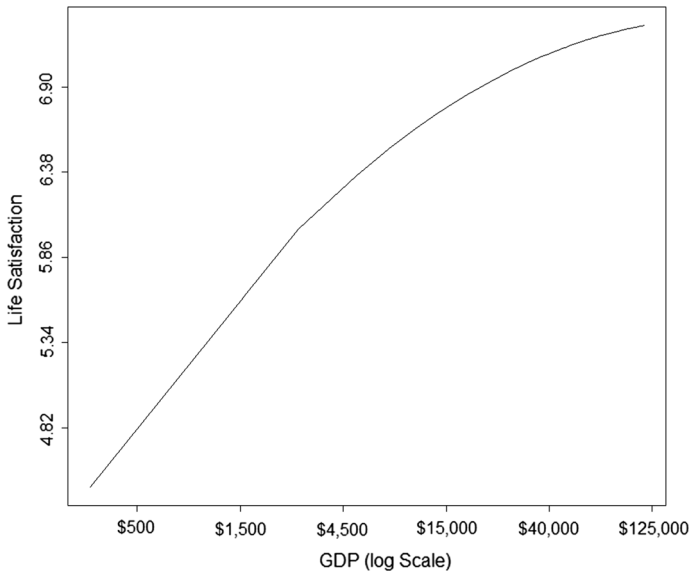
Fig. 3 Corrected relationship between GDP and life satisfaction

scales (e.g., a 3-point scale into a 10-point scale), and deflate the scales when transforming larger scales into smaller scales (e.g., an 11-point scale into a 10-point scale). While some minor differences emerged between happiness and life satisfaction (the lack of significance for the 5 and 7-point life satisfaction scales), in general, a similar pattern of effects were found for both SWB outcomes. These results suggest that the differing scale points may introduce similar biases (in both direction and magnitude), regardless of which SWB outcome is being measured. We further assessed the impact of these biases by using GDP data and found that the scale biases may substantially impact the results of analyses.

Our findings bring two important issues to the fore. First, our findings help to clarify the impact of scale transformations using the linear stretch method; specifically that both happiness scales and life satisfaction assessed with differing scale points do not appear to be comparable following simple scale transformations. This suggests that transforming either scale for comparisons across countries or time should be carefully considered as it may introduce bias into the data. This supports the need to use consistent measures of SWB for making comparisons across nations and over time and/or to utilize other transformation methods as suggested by de Jonge et al. (2014).

Second, the somewhat distinctive impact scale transformations have on happiness and life satisfaction measures suggest a potentially more complex interaction beyond just the statistical process of scale transformation. This difference may be a result of individuals using different cognitive processes when answering questions related to happiness versus life satisfaction. For happiness measures, participants may be more focused on their momentary feelings or they may pay greater attention to scale anchors, both of which may independently impact the results (Schwarz 1999; Schwarz and Strack 1999). This difference in cognitive processing may be further exacerbated by the statistical transformation of these scales. Our study's findings should encourage additional work on the psychometrics of SWB scales for global comparisons (e.g., Diener et al. 2013).

Third, it is important to consider the implications of using scale transformed scores when examining the substantive relationship between income and happiness (e.g.,

Veenhoven and Vergunst 2014). The ongoing controversy of whether income leads to happiness over time may in part be due to unreliability introduced from the use of different scale types. One possibility of addressing this issue in the future is to model the different types of scales in the analyses to account for bias that may be introduced from scale types.

One limitation of our current study is that we focused on only two well-being dimensions (happiness and life satisfaction), so we cannot be certain that this finding would generalize across other well-being scales. Nevertheless, both the affective (happiness) and evaluative (life satisfaction) components are generally viewed as integral aspects of well-being as understood within the context of subjective well-being (Diener 1984).

Based on the present study and the suggestion of a reviewer, it will be important for future work in this area to address questions regarding the intrinsic problems and qualities of different types of scales. Understanding the innate positive and negative qualities of these scales can help to determine the types and wordings of scales that are inherently better than others as well as what the appropriate application of these scales are. Further probing of this question will be important for both secondary analyses of archival datasets as well as designing future questionnaires for interview based projects (e.g., Gallup). In addition, future work should also examine the impact of different types of scale transformations apart from the linear stretch method; however, these comparisons will depend on the ability to obtain individual-level data.

6 Conclusion

In conclusion, the results of this study suggest that researchers should proceed with caution when using transformed SWB scores for comparisons across nations or over time. This is because transformations appear to introduce systematic bias to SWB scores. In the future, aside from using uniform SWB scales for national comparisons, other more sophisticated methods need to be considered to improve scale transformation methods (e.g., de Jonge et al. 2014).

References

Abdel-Khalek, A. M. (2006). Measuring happiness with a single-item scale. *Social Behavior and Personality: An International Journal, 34*(2), 139–150.

Arampatzi, E., Burger, M. J., & Veenhoven, R. (2015). Financial distress and happiness of employees in times of economic crisis. *Applied Economic Letters, 22*(3), 173–179.

de Jonge, T., Veenhoven, R., & Arends, L. (2014). Homogenizing responses to different survey questions on the same topic: Proposal of a scale homogenization method using a reference distribution. *Social Indicators Research, 117*(1), 275–300.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542–575.

Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55*(1), 34–43.

Diener, E., & Biswas-Diener, R. (2002). Will money increase subjective well-being? A literature review and guide to needed research. *Social Indicators Research, 57*(2), 119–169.

Diener, E., & Seligman, M. E. P. (2004). Beyond Money. *Psychological Science in the Public Interest (Wiley-Blackwell), 5*(1), 1–31.

Diener, E., Inglehart, R., & Tay, L. (2013a). Theory and validity of life satisfaction scales. *Social Indicators Research, 112*(3), 497–527.

Diener, E., Tay, L., & Oishi, S. (2013b). Rising income and the subjective well-being of nations. *Journal of Personality and Social Psychology, 104*(2), 267–276.

Easterlin, R. A. (1974). Does economic growth improve the human lot? In P. A. David & M. W. Reder (Eds.), *Nations and households in economic growth* (pp. 89–125). New York: Academic Press.

Easterlin, R. A. (1995). Will raising the incomes of all increase the happiness of all? *Journal of Economic Behavior & Organization, 27*(1), 35–47.

Easterlin, Richard. A. (2005). Feeding the illusion of growth and happiness: A reply to Hagerty and Veenhoven. *Social Indicators Research, 74*(3), 429–443.

Extremera, N., & Fernandez-Berrocal, P. (2013). The subjective happiness scale: Translation and preliminary psychometric evaluation of a Spanish version. *Social Indicators Research, 119*(1), 473–481.

Fischer, R., & Boer, D. (2011). What is more important for national well-being: money or autonomy? A meta-analysis of well-being, burnout, and anxiety across 63 societies. *Journal of Personality and Social Psychology, 101*(1), 164.

Fischer, R., & Van de Vliert, E. (2011). Does climate undermine subjective well-being? A 58-nation study. *Personality and Social Psychology Bulletin, 37*(8), 1031–1041.

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature, 40*(2), 402–435.

Graham, C. (2005). Insights on the development from the economics of happiness. *World Economics, 6*(3), 41–55.

Hagerty, M. R., & Veenhoven, R. (2003). Wealth and happiness revisited- growing national income does go with greater happiness. *Social Indicator Research, 64*(1), 1–27.

Inglehart, R., Foa, R., Peterson, C., & Welzel, C. (2008). Development, freedom and rising happiness: A global perspective 1981–2006. *Perspectives on Psychological Science, 3*(4), 264–285.

Judge, T. A., & Kammeyer-Mueller, J. D. (2011). Happiness as a societal measure. *Academy of Management Perspectives, 25*(1), 30–41.

Kalmijn, W., Arends, L., & Veenhoven, R. (2010). Happiness scale interval study. Methodological considerations. *Social Indicators Research, 102*(3), 497–515.

Liao, P. (2013). More happy or less unhappy? Comparison of the balanced and unbalanced designs for the response scale of general happiness. *Journal of Happiness Studies, 15*(6), 1407–1423.

Lim, H. E. (2008). The use of different happiness rating scales: Bias and comparison problem? *Social Indicators Research, 87*(2), 259–267.

National Research Council. (2013). *Subjective well-being: Measuring happiness, suffering, and other dimensions of experience*. Washington: The National Academies Press.

OECD. (2013). *How's life? 2013: Measuring well-being*. Paris: OECD Publishing.

Ovaska, T., & Takashima, R. (2010). Does a rising tide lift all the boats? Explaining the national inequality of happiness. *Journal of Economic Interest, 44*(1), 205–223.

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-Plus*. Berlin: Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2015). Nlme: Linear and nonlinear mixed effects models. R package version 3.1-119.

Development Core Team R. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 5*(2), 93–105.

Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 61–84). New York: Russell Sage Foundation.

Steel, P., & Ones, D. S. (2002). Personality and happiness: A national-level analysis. *Journal of Personality and Social Psychology, 83*(3), 767.

Tay, L., Herian, M. N., & Diener, E. (2014). Detrimental effects of corruption on subjective well-being: Whether, how, and when. *Social Psychological and Personality Science, 6*(1), 751–759.

Veenhoven, R. (1995). World database of happiness. *Social Indicators Research, 34*(3), 299–313.

Veenhoven, R. (2004). *World database of happiness: Continuous register of research on subjective appreciation of life*. Retrieved from http://hdl.handle.net/1765/8683.

Veenhoven, R. (2009). World database of happiness tool for dealing with the 'Data-Deluge'. *Psihologijske teme, 18*(2), 221–246.

Veenhoven, R. (2014). *World Database of Happiness*, [Data Set]. Retrieved from http://worlddatabaseofhappiness.eur.nl.

Veenhoven, R., & Vergunst, F. (2014). The Easterlin illusion: Economic growth does go with greater happiness. *International Journal of Happiness and Development, 1*(4), 311–343.

World Bank (2014). GDP per capita (current US$). Retrieved from http://databank.worldbank.org/data/views/reports/tableview.aspx.