



# Measuring Happiness and Life Satisfaction amongst Swedish Citizens: an Inquiry into Semantic Equivalence in Comparative Survey Research

Sofia Axelsson<sup>1</sup> · Stefan Dahlberg<sup>2,3</sup>

Accepted: 28 September 2024  
© The Author(s) 2024

## Abstract

A consistent empirical finding is that Scandinavian countries by international standards score steadily high in terms of subjectively reported levels of happiness and life satisfaction. Intrigued by previous findings in Denmark (Lolle and Goul Anderson in *Metode Og Forskningsdesign* 1:95–119, 2013, in *Journal of Happiness Studies* 6:1–14), this paper confirms that this is partly due to language effects. In this paper, Sweden serves as a case study that, similar to the Danish study, seeks to determine whether it is possible to establish semantic equivalence between translated survey items. By using randomized experiments on a representative sample of Swedish citizens with fluent skills in English, we test the effects of different designs in question wordings and response scale labels implemented by international surveys. The results reveal significant differences in answers on happiness. While the mean differences are very small, the distribution of answers is substantial enough to confirm a strong semantic threshold between the English term *happy* the Swedish term *lycklig*. Hence, it requires something more to be “very happy” in Swedish than in English. Notably, language appears to have a lesser impact on the distribution of responses across language groups when using a numbered response scale with endpoint labels, indicating that a particular question design either mitigates or intensifies translational effects. Happiness, it is concluded, is not easily translated and survey practitioners should bear this caveat in mind when operationalizing the concept across countries and cultures.

**Keywords** Survey translation · Happiness · Life-satisfaction · Survey experiments · Cross-cultural surveys · Language effects

---

✉ Stefan Dahlberg  
stefan.dahlberg@miun.se

Sofia Axelsson  
sofia.axelso@gu.se

<sup>1</sup> Department of Political Science, University of Gothenburg, Gothenburg, Sweden

<sup>2</sup> Department of Humanities and Social Science, Mid Sweden University, Östersund, Sweden

<sup>3</sup> Department of Government, University of Bergen, Bergen, Norway

## 1 Introduction

Over the past decades, happiness and life satisfaction have increasingly gained in prominence across a wide range of disciplines including but not limited to those of medicine, psychology, humanities, economics, and social sciences. Nowadays, measures of happiness and satisfaction are frequently being deployed by large-scale comparative surveys also in conjunction with objective measures. Country variations in levels of happiness and life satisfaction accounted for by survey data are generally explained to be the cause of factors such as political systems and institutions (Rothstein, 2010), economic growth (Stevenson and Wolfers, 2008), social equality, and social capital (Graham, 2011). Despite well-known linguistic and cultural differences within as well as across countries, such potential effects are often treated as a secondary concern by the quantitative research community (Wierzbicka, 2004, 2011).

It is a robust finding that Scandinavian countries by international standards score consistently high in terms of subjectively reported levels of happiness and life satisfaction. Intrigued by previous findings in Denmark, presented by Lolle and Goul Andersen (2013, 2016), this paper argues that this is partly due to language effects. Systematic country comparisons call for measures that are not only valid and reliable but, as Lolle and Goul Andersen note, equivalent across countries. Calling attention to the criterion of equivalence, cross-national comparative studies share a major methodological challenge: the fact that surveys require translations into country-specific languages. Though many researchers are aware of this problem, raising questions of how well attitudinal terms such as “happiness” and “satisfaction” travel between languages (Duncan, 2005; Harkness and Schoua-Glusberg, 1998), few systematic studies have been undertaken.

For this paper, Sweden serves as a case study that, similar to the Danish study, seeks to determine whether it is possible to establish semantic equivalence between translated survey items. By using randomized experiments on a representative sample of Swedish citizens with fluent skills in English, we test the effects of different designs in question wordings and response scale labels implemented by international surveys. The paper is structured as follows: first, we introduce the concept of language effects in cross-cultural survey research and discuss the role of language and culture in the body of survey-based well-being research, after which we present the findings from Denmark in greater detail. From there, we dive into our Swedish case study, laying out our hypotheses and experimental setup and results before adding nuance to our findings in the final discussion.

## 2 Language Effects in Cross-Cultural Survey Research

Since the early 1980s, the number of country comparative surveys has increased substantially. Together, large-scale cross-national surveys such as the World Values Survey (WVS), the European Social Survey (ESS) and the International Social Survey Programme (ISSP), to name a few, cover a majority of the world's population. Such surveys have generated ground-breaking research on a vast array of matters central to social sciences in that they provide necessary empirical tools for the systematic study of political values, attitudes, skills, and activities of ordinary citizens living under different political circumstances and in various economic and social contexts (Jowell, 1998). Nevertheless, the benefits of cross-national survey research rest on a crucial assumption: that we are able to

make comparisons between the findings obtained in different cultural settings across countries (Braun and Harkness, 2005).

Indeed, all forms of survey research are subject to potential problems and different sources of error that may hamper country comparisons. Most often, the discussion about methodological challenges focuses on problems related to sampling and between-country differences in practices of survey design and implementation. While such problems can often be mitigated—for example through weighting adjustments or by applying standard error corrections—cross-national survey research faces another methodological challenge that is harder to circumvent; namely that of semantic equivalence. In other words, the use of cross-national survey research relies on the notion that all languages share concepts that are semantically comparable (Behr, 2023; Lolle and Goul Andersen, 2013, 2016).

Ever since Sapir (1929) and Whorf (1956) introduced the notion of linguistic relativity, many studies have provided evidence to support the argument that language affects the way we think. Linguistic components such as syntactic and grammatical structures (see Biere and Lanktree, 1983; Bickel, 1997) or semantic categories (see Hunt and Agnoli, 1991) have been shown to influence cognition across languages (Peytcheva, 2008). What we refer to as language effects in cross-cultural survey research occur when the language of administration affects the ways in which respondents—particularly bilingual bicultural individuals—answer survey questions (Pérez, 2009, 2011; Peytcheva, 2019, 2020). As Lee and colleagues (2020) put it: “the differences in responses across languages may affect not only true differences in the concept that a question seeks to measure but also measurement artifacts due to translation (2020:76). Different languages are spoken in different cultural settings and are inevitably linked to different cultural systems of meaning (Geertz, 1993). Language, according to recent psycholinguistic research, can affect the way our mind operates, not least by activating cognitive frames associated with particular cultural meaning systems (see Haberstroh et al., 2002; Schwarz, 2003). The language of administration—and the use of certain translated survey items—can therefore prime cultural-specific frames in respondents, influencing their cognitions of the questions asked and ultimately their answers to them.

Needless to say, language effects may bear significant implications for country comparative research. The research community is not unaware of this problem and the question of how well translated terms travel across languages is sometimes raised by scholars engaging with political attitudes and behaviors (Dorn et al., 2007, 2008). Yet, the extent to which political scientists address the potential effects of language vary substantially (King et al., 2004). Questions about happiness and life satisfaction are today widely implemented survey items, both over time and across countries, and with a few exceptions (see Goddard and Wierzbicka, 2014; Wierzbicka, 2004, 2011), the critical issue of language has not been systematically studied. The next section sums up the research, mainly confined within happiness studies, which have elaborated on the potential implications of language and culture.

### 3 Language Effects in Happiness Studies

The concept of happiness has long occupied the minds of philosophers and historical theorists. While the interest in happiness and place in human life has been relatively stable over time, the ways in which happiness has been understood throughout history vary substantially (Nilsson, 2012). In the words of Nilsson: “Rather, happiness has historically been a fluid concept capable of producing new meanings and definitions” (2012:225)—from

classical Greek representations of *Eudaimonia* and *hēdonē* to eighteenth century representations of utilitarianism (Ahmed, 2007; Brülde, 2007; Brülde and Bykvist, 2010).

In the recent decade, however, the concept has gained in salience also outside the realm of philosophy, not least, Sears et al. (2014) note, as a means to manage populations, optimize productivity, and maintain and improve the longevity and quality of individuals' lives. Happiness studies has developed into a field of its own, encompassing a variety of disciplines such as medicine, psychology, economics, humanities and social sciences (Kullenberg and Nelhans, 2015; Youssef and Diab, 2021). Bodies of literature on topics of subjective well-being, positive affect, satisfaction with life and individual happiness are continuously growing alongside popular and scientific international rankings and indices of the levels of happiness and well-being across populations (Bonniwell, 2017; Musikanski et al., 2017; Zwolinsky, 2019). Around the world, citizens' self-reported levels of overall happiness, satisfaction with life and with other institutional arrangements are used as proxies for measuring the quality of governments and institutions as for the success of states, systems and societies (Jovanović et al., 2019; Li and An, 2020; Rothstein and Holmberg, 2015).

Although happiness and life satisfaction are related, they tend to pertain to different aspects of well-being (Raudenská, 2023), with the former more associated with emotional quality of life and the latter more to general evaluation of life (Kahneman and Deaton, 2010). However, as subjectively reported levels of both tend to correlate with other political, economic and social indicators of well-being, scholars building on cross-national surveys generally argue that there cannot be systematic measurement errors in citizens' reporting of happiness and life satisfaction (Inglehart et al., 2008). According to Diener and colleagues (2012), errors causing variations in response patterns are likely to be related to the design and implementation of surveys—rather than semantics—but such errors can be largely identified and isolated. Put differently, consistency in the conditions and consequences of well-being, it is argued, prove that survey questions about happiness and satisfaction capture important aspects of individuals' lives. Concerns over the validity of single-item questions about life satisfaction and happiness in different countries have, however, been raised in the survey community due the complicated nature of these social and psychological phenomena (Bowling, 2005). Studies comparing such single-item questions with multiple-item questions and scalars, e.g. the Satisfaction with Life Scale (SWLS, Diener et al., 1985), have rendered mixed results with multi-item questions coming out on top in many cases, but not all (see for instance Jovanović and Lazić, 2020). Recent studies have for instance shown that while the SWLS is generally more stable and less prone to measurement error compared to single-item questions about life satisfaction (Jovanović and Lazić, 2020), it also exhibits significant measurement invariance across cultural groups (Emerson et al., 2017).

According to the more critical voices, measuring and comparing levels of happiness across countries is methodologically precarious given that attitudes, behaviors, and emotions are closely aligned with specific cultural norms and customs (He et al., 2021; MacIntyre, 1971; Oishi, 2010). Others have taken such investigations further, reaching the conclusion that functional equivalence of happiness is difficult to achieve for linguistic and cultural reasons (Wierzbicka, 2004, 2011). Some languages lack a word for happiness altogether, and even when suitable comparable concepts for happiness can be identified, these could have very different historical origins, associations, and connotations in different cultural contexts—ranging from fortune and good luck to the fulfillment of one's desires (Goddard and Wierzbicka, 2014; Oishi, 2010; Wierzbicka, 2004). Though some languages hold several words corresponding to the word happiness, the context may largely

determine how to semantically conceive of one word or another (Levisen, 2014). Local semantics, it is argued, may therefore have an impact on how respondents report on subjectively assessed attitudes and emotions given that the corresponding words are imbued with different meanings across time and place.

The correlates between subjectively reported happiness and life satisfaction have also shown to differ between individualist and collectivist cultures. According to Suh et al. (1998), social norms—conceived as social approval of life satisfaction—were found to be more important determinants of life satisfaction in collectivist cultures than in individualist cultures. In the case of individualist cultures, then, emotions tended to play a far more superior role, suggesting that the receptiveness to attitudinal and emotional concepts depend on the cultural context. In some cultures, people may be less comfortable to declare that they are either “very” happy or satisfied (Villar, 2009); in others, social expectations may instead drive respondents to over-report certain attitudes and emotions (Lolle & Goul Andersen, 2016). Individual happiness or self-estimated satisfaction with life might, furthermore, not hold the same importance in all cultures (Ahuvia, 2002; Suh et al., 1998). In contrast to individualist societies, Ahuvia (2002:31) notes, “members of collectivist cultures prioritize honor, face, and meeting their social obligations above their own happiness”.

In sum, cross-national studies on attitudes and emotions can be affected by various linguistic and cultural factors including the lack of semantically corresponding terms, local semantics, social norms, and expectations in different cultural settings, as well as the importance certain concepts might hold (or not hold) across cultures. Thus, even as several studies conducted both across societies and within multilingual contexts provide evidence that language effects play a limited role in explaining observed variations in levels of happiness (Veenhoven, 2008; 2009), there are still sufficient reasons to believe that effects stemming from language or from survey translations could account for some country differences (Harkness et al., 2004).

In addition, we know that survey design and implementation may have significant impact on the response pattern of survey respondents, particularly with regard to translation practices (Behr, 2023; Harkness, 1999). According to Villar (2009), methodological issues in cross-national surveys may arise when the sentimental loadings differ between translated concepts. Examining the effects of adding and modifying response scale points and labels, the author finds that modifications constitute a source of variability in response patterns. Substantial differences in translations of response scales thus affect extreme response styles, that is, the tendency to select the endpoints of response scales measuring various attitudes. This point is further highlighted by Lolle and Goul Andersen (2013, 2016)—albeit in a different manner—whose experimental study on language effects in Denmark serves as a backdrop for the study presented herein.

#### 4 Is Denmark Really the Happiest Nation in the World?

Denmark’s top position as one of the “happiest nations in the world” is a topic of much debate in policy and research circles alike. According to Lolle and Goul Andersen (2013, 2016), however, subjectively reported levels of happiness in Denmark are, contrary to popular belief, relatively inconsistent across time, especially so when compared to reported levels of life satisfaction amongst Danish citizens. In fact, the authors argue,

it is primarily due to Denmark's consistently high levels of life satisfaction over time that the country has attracted attention.

Conducting survey experiments with Danish students, Lolle and Goul Andersen (2013, 2016) test different versions of translated questions from a variety of international survey programs asking about individual happiness and general life satisfaction. Dividing respondents into different treatment groups, with parts of the sample answering questions in English from the master questionnaires and other parts answering translated questions from the Danish questionnaires, the experiment reveals significant differences between the two languages. The semantic threshold for reporting life satisfaction is lower in Danish than in English; when Danish questions with the translation *tilfreds* is utilized, respondents score significantly higher compared to when the English questions with the original term *satisfied* is used.

In contrast to life satisfaction, respondents provided with Danish questions about happiness, where the Danish translation *lykkelig* is utilized, report significantly lower levels of happiness compared to respondents provided with English questions with the original term *happy*. As such, Lolle and Goul Andersen (2016) note a similar, albeit reversed, semantic threshold where the Danish word *lykke* "signals something bigger and more deeply felt" than the English word *happiness*. A similar finding has also been reported by López (2017), who studied the relationship between term intensity and semantic thresholds. López found that more frequently used terms, such as *happy* or *tilfreds*, tend to have a higher semantic threshold, meaning that these words are generally understood across broader contexts before reaching their full emotional or semantic impact. In contrast, less frequent terms, such as *satisfied* or *lykkelig*, may require fewer contextual cues to convey a strong emotional or intense meaning. This supports the notion that frequent words, though common, may convey less intensity in isolation due to their broader semantic scope.

Turning to the distribution of responses across survey questions about individual happiness and life satisfaction, it is further confirmed that being "very satisfied" ("*meget tilfreds*") with life is "somewhat easier" than being "very happy" ("*meget lykkelig*") in Danish. While cautious not to draw any "firm conclusions" about the semantic variations observed in the study, the authors conclude that a similar lack of semantic equivalence between English survey questions and translated survey questions are likely to be found in other languages as well.

In line with other studies in the field (Villar, 2009, see also DeJounge et al., 2015 and Liao, 2014), Lolle and Goul Anderson call attention to the fact that response patterns and distributions of responses can suffer from measurement errors due to inadequate translations of response scale labels. Indeed, one of the questions measuring life satisfaction in their experiment displays such an error. This question is borrowed from the European Social Survey (ESS), and is the only question in the experiment that utilizes a numeric response scale ranging from 0 to 10, where the endpoint 0 is labelled "*extremely dissatisfied*" and 10 is labelled "*extremely satisfied*". The translated endpoint label *særdeles*, from the Danish ESS questionnaire, the authors note, is an inadequate substitute for the original endpoint label *extremely*, the former term being "notoriously less demanding" than the latter. The lack of semantic equivalence across the ESS question, Lolle and Goul Anderson argue, is likely the reason behind the almost negligible differences in average scores between the Danish and English treatment groups. The Danish endpoint label *særdeles* is of lesser semantic strength than the English endpoint label *extremely*, and the language effects observed when testing the item life satisfaction using other questions from the

European Values Study (EVS) and the International Social Survey Programme (ISSP) become less pronounced.<sup>1</sup>

Here, we would like to emphasize that the ESS question measuring life satisfaction is of a different design than the other survey questions tested in Lolle and Goul Anderson's survey experiment. The additional survey questions testing language effects across items measuring happiness and life satisfaction are borrowed from the EuroBarometer, the European Values Study (EVS), the International Social Survey Programme (ISSP), and all survey questions are equipped with a 4-point unipolar response scale. In addition to differences in direction and length of the response scales, these survey questions also provide labels for every scale point, whereas the ESS survey question only provide two labels, one for each endpoint of the scale.<sup>2</sup> Thus, given that the ESS question is the only question tested in the Danish study that contains a longer scale that is equipped with two—bipolar—labels, it is difficult to know whether the same result holds true also for other attitudinal questions with similar answer scale design. Along this line of reasoning, it would be interesting to further modify the ESS response scale to investigate whether more moderate endpoint labels yield higher scores.

In light of the language effects found in Lolle and Goul Anderson's survey experiment, it is reasonable to expect that a replication of their experiment in a Swedish context—given Sweden's similarities to Denmark in institutional, cultural, and linguistic terms—should produce comparable results, particularly regarding the survey item happiness.<sup>3</sup>

## 5 Hypotheses

Intrigued by the findings in Denmark, this paper sets out to investigate the extent to which similar translation discrepancies—in terms of differences in word intensity—can be found also in Sweden, a neighboring country sharing similar linguistic, cultural, and institutional history. In terms of happiness, Sweden does not enjoy a privileged status

---

<sup>1</sup> While the average difference between the treatment groups receiving different versions of the ESS question is not substantial enough to yield statistical significance, differences in the proportion of responses in category 8 and 10 at the 11-point response scale are nonetheless highly significant (Lolle and Goul Anderson 2016). Compared to the English group, the Danish group appears to have a greater propensity to choose category 8 (where 10 is maximum). Conversely, respondents from the English group (provided with endpoint label “*extremely satisfied*”) seems to be more inclined to choose category 10 (maximum) than respondents from the Danish group (provided with endpoint label “*særdeles tilfreds*”).

<sup>2</sup> The response scales for the remaining survey questions are thus equipped with labels that in English read very satisfied, fairly satisfied, not very satisfied, not at all satisfied (wordings from EuroBarometer), very happy, rather happy, not very happy, not at all happy (wordings from EVS and ISSP. For the ISSP, a fifth can't choose option was also provided). These scales do not provide any antonym options (happy versus unhappy, satisfied versus dissatisfied) and are by definition unipolar.

<sup>3</sup> This expectation is further supported by research in computational linguistics, which demonstrates a negative correlation between word frequency and word intensity (Eisenstein, 2017; López, 2017). Such findings suggest that more intense words, which are typically less frequent and longer, may evoke stronger emotional responses in survey participants. As a result, linguistic variations in word intensity, rather than introducing additional considerations, naturally align with the way people interpret survey items like happiness. For instance, given that longer and less frequent words are perceived as more intense, we might anticipate subtle differences in the emotional weight assigned to terms across languages, which could, in turn, influence survey outcomes (Lewis et al., 2014; Bennett and Goodman, 2018). Therefore, we expect that in a comparison between English and Swedish, the term happy would produce more pronounced differences than satisfied, aligning with the varying intensities associated with these terms (see Fig. 2 in the appendix for relative frequencies across languages). This linguistic insight directly supports the expected results from the survey experiment, reinforcing the idea that language shapes emotional expression in meaningful ways.

as Denmark; though Sweden is generally located in amongst the top 10 highest ranked countries in terms of happiness (see for instance Helliwell et al., 2017), it is in less stiff competition for the top 1 position and has not attracted the same international attention as its neighbor Denmark.

Yet, given that the Swedish survey translation of happy—*lycklig*—is closely related to the Danish term *lykkelig*—both sharing the same etymological roots that can be traced as far back as to Old Norse—we expect to find a similar semantic gap between English and Swedish questions about happiness as found in Denmark. In other words, it should be easier for respondents to report higher levels of happiness when asked about it in English. This, we believe, should hold true both for aggregated levels of happiness as well as for level distributions of happiness; respondents answering Swedish questions will thus have a lower probability of reporting that they are “very happy” with their lives than had they answered English questions. Along this way of reasoning, we formulate the following hypotheses:

**H1a** The Swedish term *lycklig* is more positively loaded than the English term *happy*, indicating a semantic gap between the two words. The Swedish questions about happiness will yield significantly lower average scores than the English questions.

**H1b** The distribution of responses between Swedish and English questions about individual happiness are significantly different in that the probability of choosing the category “very happy”—compared to the category “rather happy”—is higher among respondents provided with English questions.

In contrast, the Swedish survey translation of satisfied—*nöjd*—lacks the same semantic and etymological relation to the Danish translation *tilfreds* as seen in the previous comparison between *lycklig* and *lykkelig*. Additionally, it is known that Danish citizens tend to over-report levels of life satisfaction due to discrepancies in functional equivalence between the Danish and the master survey questions measuring life satisfaction. As a result, we expect fewer discrepancies between the English and Swedish questions about general life satisfaction, given the semantic differences in the translation of the term. This expectation is further supported by the relationship between term frequency and word intensity (López, 2017), where the Swedish term *nöjd* has a similar relative frequency to the English *satisfied* but differs from the Danish *tilfreds* (see Fig. A2 in the appendix). As such, we trust that the Swedish translation more closely aligns with its English equivalent than does the Danish translation. This leads us to our next hypothesis:

**H2a** The Swedish term *nöjd* is neither more nor less—positively or negatively—loaded than the English term *satisfied*, indicating the absence of a semantic gap between the two words. The Swedish questions about life satisfaction will not yield average scores that are significantly different from the scores of the English questions.

Provided that we hypothesize a (more or less) semantic agreement between English and Swedish in questions about life satisfaction, we do not expect any significant differences in level distribution of life satisfaction across the two languages. Put differently, language per se is not believed to be affecting the probability of respondents to favor any particular response category. Our final hypothesis thus states:

**H2b** The distribution of responses between Swedish and English questions about general life satisfaction are not significantly different; respondents' probability of choosing one response category before another is not an effect of the language of administration of the survey question.

We test this hypothesis primarily by replicating the Danish survey experiment, this time with a large sample of Swedish citizens.

## 6 Data, Design and Measurements

For this experiment, we have attempted an as close to as possible replica of the Danish survey experiment; thereby using the same English survey questions but swapping the Danish translated questions for the Swedish. The questions, taken from the EuroBarometer, the European Social Survey (ESS), the European Values Survey (EVS), and the International Social Survey Programme (ISSP) are presented in Table 1 below. In total, we have used four treatment groups, each of which received one question asking about general life satisfaction—either from the EuroBarometer or from the ESS—and one question about individual happiness—either from the EVS or the ISSP. Treatment groups 1 and 2 were provided with the master questionnaire versions of the questions in English, while groups 3 and 4 received the translated questionnaire versions in Swedish. Moreover, groups 1 (English) and 3 (Swedish) received questions from the same waves of the Eurobarometer and the EVS, while groups 2 (English) and 4 (Swedish) received questions from the ESS and the ISSP. Some errors occur in the original design of the survey questions from the EuroBarometer and the ISSP, which was addressed by adding one control group to treatment group 3 with respect to the life satisfaction question and another one to treatment group 4 with respect to the happiness question (see Tables 5 and 6 in the appendix for the survey design including the control groups and randomization check).

The empirical data was collected by the Laboratory of Opinion Research (LORE), at the University of Gothenburg in May 2016, and is a sample—stratified by age, gender, and education—from the Swedish Citizen Panel, which is an online respondent panel administered by the University of Gothenburg. In the randomized process of creating treatment groups, some respondents were given the opportunity to answer questions in English. It should be noted that English is not an official language in Sweden but the level of English proficiency amongst Swedish citizens is very high by international standards.<sup>4</sup> Against this backdrop, we argue that there are sufficient reasons to trust the results as valid, even though we should be careful to draw any definite conclusions. Randomization into treatment groups was made based on a screening question asking whether participants were willing to answer survey questions in English as a part of an international study. However, to control for language skills was not possible in any of the survey experiments, something that ultimately poses a limitation of the experiment. The participation rate was 64 percent of the entire gross sample, generating a total of approximately 3,500 participants that were randomized into four different treatment groups.

<sup>4</sup> The EF English Proficiency Index ranks Sweden as top 2 out of 70 countries in Europe. With a score of 70.40, Sweden is considered to have very high English proficiency compared to the European average score 55.69, which is considered moderate.

**Table 1** Replication experiment: survey design

English treatment groups	Swedish treatment groups
Life satisfaction: treatment group 1 On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead? Response categories: very satisfied, Fairly satisfied, Not very satisfied, Not at all satisfied Wording: EuroBarometer 2006	Life satisfaction: treatment group 3 På det stora hela, är du mycket nöjd, ganska nöjd, ganska missnöjd eller mycket missnöjd med det liv du lever? Skulle du säga att du är... Response categories: Mycket nöjd, Ganska nöjd, Ganska missnöjd, Mycket missnöjd Wording: EuroBarometer 2006
Life satisfaction: treatment group 2 All things considered, how satisfied are you with your life as a whole nowadays? Response categories: 0–1 where 0 is labelled Extremely dissatisfied and 10 is labelled Extremely satisfied Wording: ESS 2010	Life satisfaction: treatment group 4 På det hela taget, hur nöjd är du med ditt liv i stort nuförtiden? Response categories: 0–1 where 0 is labelled Extremt missnöjd and 10 is labelled Extremt nöjd Wording: ESS 2010
Happiness: treatment group 1 Taking all things together, would you say you are: Response categories: very happy, Rather happy, Not very happy, Not at all happy Wording: EVS 1990	Happiness: treatment group 3 Skulle du, allmänt sett, beskriva sig själv som: Response categories: Mycket lycklig, Ganska lycklig, Inte särskilt lycklig, Inte alls lycklig Wording: EVS 1990
Happiness: treatment group 2 If you were to consider your life in general these days, how happy or unhappy would you say you are, on the whole? Response categories: very happy, Fairly happy, Not very happy, Not at all happy Wording: ISSP 2007	Happiness: treatment group 4 Om du betraktar ditt liv i största allmänhet, hur lycklig eller olycklig skulle du säga att du på det hela taget är? Response categories: Mycket lycklig, ganska lycklig, Inte särskilt lycklig, Inte alls lycklig Wording: ISSP 2007

For treatment group 3 (life satisfaction), the question and response scale are of bipolar design. We therefore added an additional control group with a revised version of the Swedish EuroBarometer question. For treatment group 4 (happiness), the question is of bipolar design but with a unipolar response scale. We therefore added an additional control group with a revised version of the Swedish ISSP question. More information about the control groups can be found in the appendix

In addition to the replication experiment presented above, we conducted another survey experiment using one question about happiness and one about life satisfaction with 11-point response scales provided by the ESS.<sup>5</sup> While this other experiment also tested the effect of language on the distribution of responses and average response scores, we are here interested in comparing fully labeled ordinal scales with longer rating scales only provided with end-point labels.

This experiment (hereafter referred to as the response scale experiment) has the advantage that it allows us to more thoroughly compare responses of the survey item life

<sup>5</sup> Data for this other experiment was collected in March 2016, another stratified sample from the Swedish Citizen Panel. The sample consisted of eight different treatment groups—two English groups and six Swedish groups created through the same randomization process—with approximately 3500 respondents in total. More information about the treatment groups is available in Table 7 and 8 in the appendix.

**Table 2** Replication experiment: summary statistics of survey item happiness

Group	Language	Obs	Mean	SD	Min	Max
1	English	563	3.15	0.70	1	4
2	English	544	3.14	0.70	1	4
3	Swedish	1082	3.10	0.64	1	4
4	Swedish	584	3.04	0.68	1	4
4_c	Swedish	567	3.05	0.63	1	4

Control group 3 received the same survey question about happiness as treatment group 3. Likewise, control group 4 received the same life satisfaction question as group 4. As a result, the statistics for these groups have been combined, hence the large number of observations. As no significant differences between treatment group 3 and control group 3 as well as between group 4 and control group 4 were found, these groups have been combined in the following analyses

satisfaction with those of the survey item happiness using the same numeric response scale.<sup>6</sup> For information on the ESS survey questions and the survey design of the response label experiment, see Tables 7 and 8 in the appendix. Since all but one question in our replication experiment (see Table 1 again) utilizes fully labelled 4-point response scales, we argue that a specific response scale experiment of this kind will unveil additional information on response patterns within and across languages.

## 7 Replicating the Danish Experiment

If the hypothesis that the Swedish term *lycklig* is more strongly positively loaded than the English term *happy* (Hypothesis 1a) holds true, it should mean that respondents provided with Swedish questions should report lower scores than those provided with English questions. Table 2 presents summary statistics for the survey item happiness across all treatment groups including the control groups. In total, we have five treatment groups for the survey item happiness, which all have comparable (4-point and fully labelled) response scales. The mean differences between the groups are very small in size, albeit pointing in the hypothesized direction with slightly higher mean scores (0.04–0.11) for the English groups.

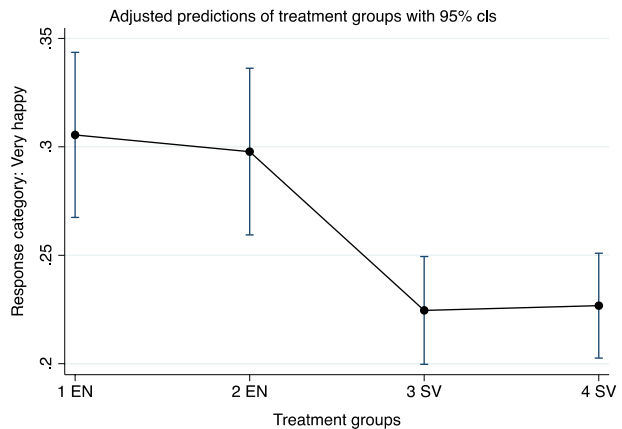
The five treatment groups provide us with nine possible two-sample comparisons; one English-English comparison (groups 1 and 2), six English-Swedish comparisons (groups 1 and 3, 1 and 4, 1 and 4\_c, 2 and 3, 2 and 4, and 2 and 4\_c) and two Swedish-Swedish comparisons (groups 3 and 4, and 3 and 4\_c). T-tests for differences in mean scores reveal significant differences at the 95 percent confidence level between six of eight two-samples—all English-Swedish comparisons. Notably, neither the English-English two-sample comparison, nor the Swedish-Swedish comparisons yield significantly different scores, indicating that something interesting is at play between the two languages.

<sup>6</sup> The replication experiment only tests one survey item from the ESS, the life satisfaction item, which is provided to treatment groups 2 (English question) and 4 (Swedish question). This complicates within-group comparisons with respect to the item life satisfaction and the item happiness for groups 2 and 4 because the response scales measuring the two items differ.

**Table 3** Replication experiment: distribution of response categories across language groups

Response category	English treatment groups (%)	Swedish treatment groups (%)
Very happy	30	23
Rather/Fairly happy	56	62
Not very happy	12	14
Not at all happy	2	1
	100	100

The table contains the distribution of responses between language groups including treatment group 1 (English EVS question), group 2 (English ISSP question), group 3 (Swedish EVS question), group 4 (Swedish ISSP question) and group 4\_c (Swedish ISSP control question)

**Fig. 1** Replication experiment: Adjusted predictions of treatment groups in choosing response category “very happy”. *Note:* Group 1 EN represents the English EVS question and group 3 SV represents the Swedish EVS question. Group 2 EN represents the English ISSP question and group 4 SV represents the Swedish ISSP question including the control group

Looking more specifically at the distribution of responses in levels of happiness (Hypothesis 1b) in Table 3—showing the proportion of responses across the two language groups for response categories 3 and 4—a clearer picture emerges. The differences between response category 3 and response category 4 is 26 percentage points for the English treatment groups and 39 percentage points for the Swedish groups, with a relative difference of 13 percentage points between the language groups. Thus, as hypothesized, it appears as if the respondents provided with Swedish questions are more inclined to choose the third response category “rather/fairly happy” over the fourth category “very happy”, compared to those provided with English questions.

Since all happiness questions are categorized across a 4-point scale, we employ a multinomial regression to predict the probability of respondents to choose category 4—“very happy”—when category 3 is used as a baseline. We do so by testing the survey questions about happiness from the EVS and the ISSP across the four different treatment groups. The regression results further confirm that the difference between categories 3 and 4 is significant at the 95 percent confidence level. As regression coefficients are somewhat difficult to interpret, the results are provided in the appendix. Instead, the adjusted predictions with margins are provided in Fig. 1.

**Table 4** Replication experiment: summary statistics of survey item life satisfaction

Group	Language	Obs	Mean	SD	Min	Max
1	English	563	3.23	0.68	1	4
2	English	546	7.10	1.83	0	10
3	Swedish	563	3.24	0.63	1	4
3_c	Swedish	520	3.20	0.65	1	4
4	Swedish	1154	7.10	1.78	0	10

Control group 3 received the same survey question about happiness as treatment group 3. Likewise, control group 4 received the same life satisfaction question as group 4. As a result, the statistics for these groups have been combined, hence the large number of observations. As no significant differences between treatment group 3 and control group 3 as well as between group 4 and control group 4 were found, these groups have been combined in the following analyses

As expected, there is a stark and significant contrast between the different groups based on language, this in spite of variations in phrasing of the question asking respondents to rate their individual happiness by the EVS vis-à-vis the ISSP. When asked about their level of individual happiness in English, 31 percent in group 1 (EVS question) and 30 percent in group 2 (ISSP question) are predicted to answer that they are “*very happy*” (category 4). In contrast, only 22 percent of respondents in group 3 (EVS question) and 23 percent in group 4 (ISSP question) are predicted to opt for this category. Instead, they have a much larger (and significant) propensity to answer that they are “*rather/fairly happy*” (see Table 9 in the appendix). The difference—ranging between 7 and 9 percentage points—between the two language groups is substantial enough to confirm both Hypotheses 1a and 1b. Even as the mean differences between the Swedish and English groups are very small, there is a clear semantic gap between the English term *happy* and its Swedish translation *lycklig* with respondents from Swedish groups being less inclined to report that they are “*very happy*”.

Turning to the survey item measuring general satisfaction with life, we expect language to have a limited impact both on the average levels of reported life satisfaction as well as on the distribution of responses (Hypotheses 2a and 2b). In other words, no survey question should yield higher or lower scores that are significantly different from those of another question, and this should hold true across all treatment groups. Moreover, the distribution of responses across the Swedish questions should be no different from those of the English questions.

The question from the Eurobarometer and the question from the ESS differ in response scale design (the former has 4-point fully labelled scale, and the latter has an 11-point numeric scale with endpoint labels), which somewhat limits our options for comparability (see Table 4). The Eurobarometer question provides us with three possible two-sample comparisons; two English-Swedish comparisons (groups 1 and 3, and groups 1 and 3\_c) and one Swedish-Swedish comparison (groups 3 and 3\_c). The ESS question only enables one comparison, between the English group 2 and the Swedish group 4.

As expected, there are no significant differences between any of the treatment groups provided with the Eurobarometer question or the ESS question when t-tests are performed for all possible two-sample comparisons. Nor is the distribution of responses in levels of life satisfaction significantly different across any of the groups. The null results lend support to Hypotheses 2a and 2b; we do not detect a semantic gap between the term *satisfied* and its Swedish translation *nöjd*. In terms of measuring life satisfaction, it appears as though the international

survey programs have been more successful in reaching semantic equivalence in Swedish surveys compared to Danish surveys.

## 8 Adding Nuance to the Findings in Denmark and Sweden

When comparing our results with the findings of Lolle and Goul Andersen (2016), we can conclude that in the Swedish case, there is a language effect for happiness but not for life satisfaction. These findings are not surprising. The translated term for happiness is, in both cases, the etymologically identical term *lykkelig/lycklig*. If the differences between English and the translated Scandinavian terms are a function of language, we should expect similar results, which is what we observe. For life satisfaction, on the other hand, we find no significant differences between the English term *satisfied* and the translated Swedish term *nöjd*. The Danish translation for this survey item is *tilfreds*, which also has a counterpart in Swedish, though *nöjd* is a different word with different intensity and language frequency. In the Swedish case, it appears that *nöjd* is the most suitable translation for this question.

Language may very well cause unwarranted effects but, following our line of reasoning, it seems to matter more when respondents are provided with shorter answer scales containing labels of gradually increasing or decreasing semantic strength. For instance, when offering answer alternatives labelled *very satisfied*, *fairly satisfied*, *not very satisfied*, *not at all satisfied*, translations practices should be careful not to under- or overload any modifiers or sentiments. If true, this has implications not only for cross-national happiness studies but possibly for other forms of comparable research on self-reported attitudes and emotions.

Finally, we have, as mentioned in the method and measurement section, extended the Swedish experiment by also testing the translational effect for the happiness question using numerical 11-point scales (see Tables 7 and 8 in the appendix). The results from the response scale experiments we cannot detect any significant differences across languages for the numeric 11-point scales measuring life-satisfaction. Importantly, the results of the ESS question measuring life satisfaction in the Danish study reveal a rather similar lack of semantic gap, with the average score being 7.91 and 7.75 for the Danish and the English treatment groups respectively (Lolle and Goul Andersen, 2016). Albeit not statistically significant, the small difference in average scores still points in the same direction as our conclusion, the implication being that language plays a limited role for respondents when faced with a scale design similar to that of the ESS.

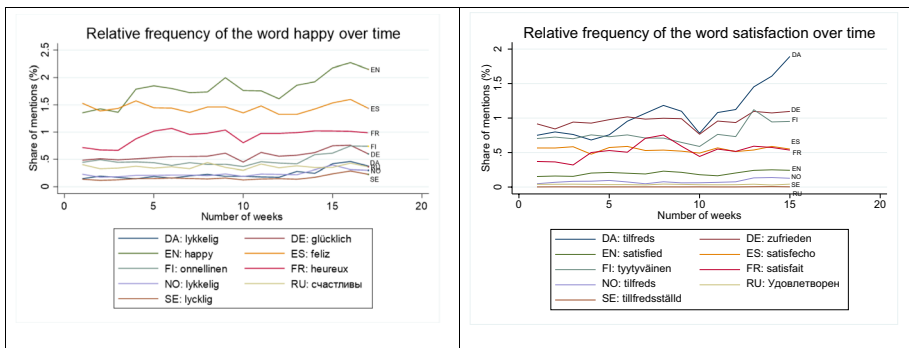
One initial takeaway from our findings is that longer numeric response scales with endpoint labels should be able to alleviate language effects of this kind. However, we also know that longer numeric scales are fraught with other types of measurement invariance, not least since they introduce more variability and longer standard errors (Lundmark et al., 2016). As such, it is not quite clear whether such response scale design still obscures existing language effects or if they, in fact, reduce such potential effects altogether. Although existing research has found that fully labelled response scales often outperform longer numeric response scales with endpoint labels, this is generally the case if the labels used are of high quality (Lundmark, 2022). If such findings hold in a cross-cultural setting, where semantic equivalence between original and translated labels cannot be fully guaranteed, remains to be seen. Hence, more experimental research is warranted to gain more insights into the nexus between response scale design and translation. While fully labelling an 11-point scale may be challenging, a seven-point scale might suffice, especially when comparing a fully labeled scale to one that only uses endpoint labels. The fact that

this study only explores four-point and 11-point scales, rather than also including five- or seven-point labelled and unlabelled scales remains a limitation and an avenue for future study.

It is evident that the concept of *lykke/lycka* in Denmark and Sweden differs slightly from the concept of happiness in English. At the same time, the widespread use of *lykke/lycka* in large-scale comparative surveys indicates that no other translations are considered more suitable. Exploring in detail the extent to which single-item questions about happiness correspond to existing multiple-item scales—such as SPANE (Diener et al., 2009)—which include different facets of happiness, would help address concerns over the different forms of validity regarding the survey questions tested here. Such explorations, inspired by recent efforts in the areas of happiness, life satisfaction, and well-being (Jovanović and Lazić, 2020; Jovanović et al., 2019; Raudenská, 2023), albeit with a stronger focus on language and local semantics, would add further nuance to our findings. We conclude that happiness is not easily translated, and survey practitioners should bear this in mind when operationalizing the concept, not only in Scandinavia, but across other countries and languages as well.

## Appendix

See Fig. 2 and Tables 5, 6, 7, 8 and 9.



**Fig. 2** Relative frequency of translational terms for happiness and life satisfaction. *Note:* The relative frequency of the words happy and satisfied, along with their respective translations in the World Value Surveys, as they occur in online text data over time. In total, the relative term frequency in nine different languages was monitored during 2016 as part of the interdisciplinary research project Linguistic Explorations of Societies (see [www.gu.se/en/linguistic-explorations-of-societies](http://www.gu.se/en/linguistic-explorations-of-societies) for more information). For further details on relative term frequency in online text data across languages, see Dahlberg et al. (2023)

**Table 5** Replication experiment: survey design including control groups

English treatment groups	Swedish treatment groups
<p>Life satisfaction: treatment group 1</p> <p>On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead?</p> <p>Response categories: very satisfied, Fairly satisfied, Not very satisfied, Not at all satisfied</p> <p>Wording: Eurobarometer 2006</p>	<p>Life satisfaction: treatment group 3</p> <p>På det stora hela, är du mycket nöjd, ganska nöjd, ganska missnöjd eller mycket missnöjd med det liv du lever? Skulle du säga att du är...</p> <p>Response categories: Mycket nöjd, Ganska nöjd, Ganska missnöjd, Mycket missnöjd</p> <p>Wording: Eurobarometer 2006</p> <p>Life satisfaction: control group 3</p> <p>På det stora hela, är du mycket nöjd, ganska nöjd, inte särskilt nöjd eller inte alls nöjd med det liv du lever? Skulle du säga att du är...</p> <p>Response categories: Mycket nöjd, Ganska nöjd, Inte särskilt nöjd, Inte alls nöjd</p> <p>Wording: revised Eurobarometer 2006</p>
<p>Life satisfaction: treatment group 2</p> <p>All things considered, how satisfied are you with your life as a whole nowadays?</p> <p>Response categories: 0–1 where 0 is labelled Extremely dissatisfied and 10 is labelled Extremely satisfied</p> <p>Wording: ESS 2010</p>	<p>Life satisfaction: treatment group 4</p> <p>På det hela taget, hur nöjd är du med ditt liv i stort nuförtiden?</p> <p>Response categories: 0–1 where 0 is labelled Extremt missnöjd and 10 is labelled Extremt nöjd</p> <p>Wording: ESS 2010</p>
<p>Happiness: treatment group 1</p> <p>Taking all things together, would you say you are:</p> <p>Response categories: very happy, Rather happy, Not very happy, Not at all happy</p> <p>Wording: EVS 1990</p>	<p>Happiness: treatment group 3</p> <p>Skulle du, allmänt sett, beskriva sig själv som:</p> <p>Response categories: Mycket lycklig, ganska lycklig, Inte särskilt lycklig, Inte alls lycklig</p> <p>Wording: EVS 1990</p>
<p>Happiness: treatment group 2</p> <p>If you were to consider your life in general these days, how happy or unhappy would you say you are, on the whole?</p> <p>Response categories: very happy, Fairly happy, Not very happy, Not at all happy</p> <p>Wording: ISSP 2007</p>	<p>Happiness: treatment group 4</p> <p>Om du betraktar ditt liv i största allmänhet, hur lycklig eller olycklig skulle du säga att du på det hela taget är?</p> <p>Response categories: Mycket lycklig, ganska lycklig, Inte särskilt lycklig, Inte alls lycklig</p> <p>Wording: ISSP 2007</p> <p>Happiness: control group 4</p> <p>Om du betraktar ditt liv i största allmänhet, hur lycklig skulle du säga att du på det hela taget är?</p> <p>Response categories: Mycket lycklig, ganska lycklig, Inte särskilt lycklig, Inte alls lycklig</p> <p>Wording: revised ISSP 2007</p>

**Table 6** Randomization control

	Group 1 (eng)	Group 2 (eng)	Group 3 (swe)	Group 4 (swe)
Sex				
Woman	45.8	47.1	44.9	45.7
Man	54.2	52.9	55.1	54.3
Age				
< 30	15.6	15.1	14.9	15.7
30–39	19.1	22.9	22.4	21.3
40–49	25.1	19.8	23.1	22.6
50–59	23.2	20	22.5	21.1
60–69	15.9	19.8	15.6	17.4
> 70	1.1	2.4	1.5	1.9
Education				
Primary school or less	2.8	3.5	4.2	3.6
High school or equivalent	24.6	26.2	29.5	29.1
Post-secondary education	16.1	15.1	15.7	14.6
College, less than 3 years	24.1	25.2	22.1	24.5
University, 3 years or longer	32.4	30	28.5	28.2
Political interest				
Very interested	43.2	44.3	41.1	41.5
Somewhat interested	49.7	47.9	52.2	51.9
Not very interested	6.7	7.6	6.1	5.9
Not at all interested	0.4	0.2	0.6	0.7
N:	565	550	1084	1158

**Table 7** Additional response label experiment: survey questions

Language	Question about happiness	Question about life satisfaction
English	Taken all things together, how happy would you say you are?	All things considered, how satisfied are you with your life as a whole nowadays?
Swedish	Allt sammantaget, hur lycklig skulle du säga att du är?	På det hela taget, hur nöjd är du med ditt liv i stort nuförtiden?

Questions from the ESS English and Swedish questionnaires with a numeric and bipolar 11-point scale including endpoint labels. For the English questions, 0 is labelled “Extremely unhappy/dissatisfied” and 10 labelled “Extremely happy/satisfied” according to the master questionnaire. For the Swedish questions, 0 is labelled “Extremt olycklig/missnöjd” and 10 is labelled “Extremt lycklig/nöjd”

**Table 8** Additional response label experiment: survey design

Treatment group		Response scale modification		Question wording	
Number	Language	Endpoint label strength		Key concept wording (happiness)	Key concept wording (life satisfaction)
<b>1</b>	<b>English</b>	<b>Extremely</b>		<b>Happy</b>	<b>Satisfied</b>
2	English	Very		Happy	Satisfied
<b>3</b>	<b>Swedish</b>	<b>Extremely</b>		<b>Lycklig</b>	<b>Nöjd</b>
4	Swedish	Very		Lycklig	Nöjd
5	Swedish	Extremely		Lycklig	Tillfreds
6	Swedish	Very		Lycklig	Tillfreds
7	Swedish	Extremely		Lycklig	Tillfredsställd
8	Swedish	Very		Lycklig	Tillfredsställd

Questions from ESS with a numeric 11-point scale and endpoint labels where 0 indicates “Extremely unhappy/dissatisfied” and 10 indicates “Extremely happy/satisfied”. Groups 1 and 3, marked in bold, have received the original question and response scale label from the English and Swedish ESS questionnaires. Groups 2 and 4 have received the original questions with modified endpoint labels that signify lesser semantic strength. Groups 5 and 7 have received modified questions—where the Swedish translation *nöjd* has been swapped with semantically similar words sometimes used in other comparative survey programs—with original endpoint labels. Groups 6 and 8 have received both modified questions as well as modified endpoint labels

**Table 9** Replication experiment: results from multinomial logistic regression

Treatment groups	Category 1 “Not at all happy”	Category 2 “Not very happy”	Category 3 “Rather/fairly happy”	Category 4 (baseline) “Very happy”
1. English (reference)				
2. English	2.275*	0.811	1.062	
	1.416	−0.17	−0.144	
3. Swedish	1.416	1.406*	1.547***	
	−0.67	−0.247	−0.187	
4. Swedish	1.506	1.593***	1.479***	
	−0.698	−0.273	−0.177	
Constant	0.0407***	0.419***	1.814***	
	−0.0157	−0.0588	−0.172	
Observations	3.340	3.340	3.340	3.340
Log likelihood	−3277	−3277	−3277	−3277
Degrees of freedom	9	9	9	9
Chi2	33.62	33.62	33.62	33.62

The table shows results from multinomial logistic regression where category 4 (“very happy”) is used as a baseline. Group 1 contains responses from the English EVS question about happiness, and is treated as a reference category. Group 2 contains responses from the English ISSP question, group 3 from the Swedish EVS question, and group 4 from the Swedish ISSP question. Significance levels are defined as follows: \*\*\*  $p < 0.01$  \*\*  $p < 0.05$  \*  $p < .1$

**Funding** Open access funding provided by Mid Sweden University. Vetenskapsrådet, 2017-02429, Stefan Dahlberg.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmed, S. (2007). The happiness turn. *New Formations*, 63(1), 7–14.
- Ahuvia, A. C. (2002). Individualism/collectivism and the cultures of happiness: A theoretical conjecture on the relationship between consumption, culture and subjective well-being at the national level. *Journal of Happiness Studies*, 3, 23–36.
- Behr, D. (2023). Translating questionnaires. In C. U. Krägeloh, M. Alyami, & O. N. Medvedev (Eds.), *International handbook of behavioral health assessment*. Springer. [https://doi.org/10.1007/978-3-030-89738-3\\_2-1](https://doi.org/10.1007/978-3-030-89738-3_2-1)
- Bennett, E., & Goodman, N. (2018). Extremely costly intensifiers are stronger than quite costly ones. *Cognition*, 178, 147–161. <https://doi.org/10.1016/j.cognition.2018.05.011>
- Bickel, B. (1997). Spatial operations in deixis, cognition, and culture: where to orient oneself in Belhare. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 46–83). Cambridge University Press. <https://doi.org/10.1017/CBO9781139086677.003>
- Boniwell I. (2017). Introduction in the centre for bhutan studies and GNH. In *Happiness: Transforming the development landscape*. Retrieved September 8, 2024 from [http://www.bhutanstudies.org.bt/wp-content/uploads/2017/05/Happiness-transform\\_Final\\_with-cover.pdf#page=44](http://www.bhutanstudies.org.bt/wp-content/uploads/2017/05/Happiness-transform_Final_with-cover.pdf#page=44).
- Bowling, A. (2005). Just one question: If one question works why ask several? Editorial. *Journal of Epidemiology and Community Health*, 59, 342–345.
- Brülde, B. (2007). Happiness theories of the good life. *Journal of Happiness Studies*, 8, 15–49.
- Briere, J. & Cheryl Lanktree. (1983). Sex-role related effects of sex bias in language. *Sex roles*, 9(5), 625–632.
- Brülde, B., & Bykvist, K. (2010). Happiness, ethics, and politics: introduction, history and conceptual framework. *Journal of Happiness Studies*, 11, 541–551.
- Dahlberg, S., Axelsson, S., Gyllensten, A. C., Sahlgren, M., Ekgren, A., Holmberg, S., & Schwarz, J. A. (2023). A distributional semantic online lexicon for linguistic explorations of societies. *Social Science Computer Review*, 41(2), 308–329. <https://doi.org/10.1177/08944393211049774>
- DeJouge, T., Veenhoven, R. & Arends, L. (2015). 'Very Happy' is Not Always Equally Happy in the Meaning of Verbal Response Options in Survey Questions. *Journal of Happiness Studies*, 16, 77–101.
- Duncan, G. (2005). *What Do We Mean by "Happiness"? The Relevance of Subjective Wellbeing to Social Policy*, *Social Policy Journal of New Zealand*, 25(25), 16–31.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49, 71–75.
- Diener, E., Inglehart, R., & Tay, L. (2012). Theory and validity of life satisfaction scales. *Social Indicators Research*, 112(3), 497–527.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-W., Oishi, S., & Biswas-Diener, R. (2009). Scale of positive and negative experience. *APA PsycTests*. <https://doi.org/10.1037/t03125-000>
- Dorn, D., Fischer, J., Kirchgässner, G., & Sousa-Poza, A. (2007). Is it culture or democracy? The impact of democracy and culture on happiness. *Social Indicators Research*, 82, 505–526.
- Dorn, D., Fischer, J., Kirchgässner, G., & Sousa-Poza, A. (2008). Direct democracy and life satisfaction revisited: New evidence for Switzerland. *Journal of Happiness Studies*, 9, 227–255.
- Eisenstein, J. (2017). Unsupervised learning for lexicon-based classification. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)* (pp. 3188–3194). AAAI Press.
- Emerson, S. D., Guhn, M. & Gadermann, A. M. (2017). *Measurement invariance of the Satisfaction with Life Scale: reviewing three decades of research*. *Quality of Life Research*, 26, 2251–2264.

- Geertz, C. (1993). *The interpretation of cultures*. Basic Books.
- Goddard, C., & Wierzbicka, A. (2014). *Words and meanings: Lexical semantics across domains, languages, and cultures*. Oxford University Press.
- Graham, C. (2011). *The pursuit of happiness: An economy of well-being*. Brookings Institution.
- Haberstroh, S., Oyserman, D., Schwarz, N., Kuhnen, U., & Ji, L. (2002). Is the Interdependent self more sensitive to question context than the independent self? Self-construal and the observation of conversational norms. *Journal of Experimental Social Psychology*, 38, 323–329.
- Harkness, J. (1999). In pursuit of quality: Issues for cross-national survey research. *International Journal of Social Research Methodology*, 2(2), 125–140.
- Harkness, J., & Schoua-Glusberg, A. (1998). Questionnaires in translation. In J. Harkness (Ed.), *Cross-cultural survey equivalence* (pp. 87–126). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbnresolving.org/urn:nbn:de:0168-ssoar-49733-1>
- Harkness, J., Pennell, B.-E., & Schoua-Glusberg, A., et al. (2004). Survey questionnaire translation and assessment. In S. Presser (Ed.), *Methods for testing and evaluating survey questionnaires*. Wiley.
- He, J., Chi, R., & van de Vijver, F. J. (2021). People use scales differently: Dealing with survey response styles in cross-cultural research. *Journal of Intercultural Communication & Interactions Research*, 1(1), 83–100.
- Helliwell, J., Layard, R., & Sachs, J. (2017). *World happiness report 2017*. Sustainable Development Solutions Network.
- Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review*, 98, 377–389.
- Inglehart, R., Foa, R., Peterson, C., & Welzel, C. (2008). Development, freedom and rising happiness: A global perspective (1981–2007). *Perspectives on Psychological Science*, 3(4), 264–295.
- Jovanović, V., Cummins, R. A., Weinberg, M., Kaliterna, L., & Prizmic-Larsen, Z. (2019). Personal wellbeing index: A cross-cultural measurement invariance study across four countries. *Journal of Happiness Studies*, 20, 759–775.
- Jovanović, V., & Lazić, M. (2020). Is longer always better? A comparison of the validity of single-item versus multiple-item measures of life satisfaction. *Applied Research in Quality of Life*, 15, 675–692.
- Jowell, R. (1998). How comparative is comparative research? *American Behavioral Scientist*, 42(2), 168–177.
- King, G., Murray, C., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 191–207.
- Kullenberg, C., & Nelhans, G. (2015). The happiness turn? Mapping the emergence of “happiness studies” using cited references. *Scientometrics*, 103, 615–630.
- Lee, S., Hu, M., Liu, M., & Kelley, J. (2020). Quantitative evaluation of response scale translation through a randomized experiment of interview language with bilingual english- and spanish-speaking latino respondents. In M. Sha & T. Gabel (Eds.), *The essential role of language in survey research*. RTI Press.
- Levisen, C. (2014). The story of “Danish Happiness” global discourse and local semantics. *International Journal of Language and Culture*, 1(2), 174–193.
- Lewis, M., Sugarman, E., & Frank, M. (2014). The structure of the lexicon reflects principles of communication. *Proceedings of the Cognitive Science Society*, 36(36), 845–850.
- Li, Q., & An, L. (2020). Corruption takes away happiness: Evidence from a cross-national study. *Journal of Happiness Studies*, 21(2), 485–504.
- Liao, P. (2014). *More Happy or Less Unhappy? Comparison of the Balanced and Unbalanced Designs for the Response Scale of General Happiness*. *Journal of Happiness Studies*, 15, 1407–1423.
- Lolle, H., & Goul Andersen, J. (2013). Er danskerne så lykkelige? Måling af lykke og tilfredshed og international surveyundersøgelser. *Metode Og Forskningsdesign*, 1, 95–119.
- Lolle, H., & Goul Andersen, J. (2016). Measuring happiness and overall life satisfaction: A danish survey experiment on the impact of language and translation problems. *Journal of Happiness Studies*, 6, 1–14.
- López, R. P. (2017). Word frequency as a predictor of word intensity. In *Master thesis in language technology*. Uppsala University.
- Lundmark, S. (2022). Verbal labels and the number of response options in grid-type questions. In *The SOM Institute's Notes on Survey Methodology—2022:2*. The SOM Institute, University of Gothenburg.
- Lundmark, S., Gilljam, M., & Dahlberg, S. (2016). Measuring generalized trust: An examination of question wording and the number of scale points. *Public Opinion Quarterly*, 80(1), 26–43. <https://doi.org/10.1093/poq/nfv042>
- MacIntyre, A. (1971). Is a science of comparative politics possible? In *Against the Self-Images of the Age, Essays on Ideology and Philosophy*. Schocken Books.

- Musikanski, L., Cloutier, S., Bejarano, E., Briggs, D., Colbert, J., Strasser, G. & Russel, S. (2017). *Happiness Index Methodology*. Journal of Social Change, 9(1), 4–31.
- Nilsson, A. (2012). *Lyckans betydelse. Sekularisering, sensibilisering och individualisering i svenska skillingtryck* (pp. 1750–1850). Agerings bokförlag.
- Oishi, S. (2010). Culture and well-being: Conceptual and Methodological issues. In E. Diener, D. Kahneman, & J. John Helliwell (Eds.), *International differences in well being*. Oxford Scholarship Online.
- Pérez, E. O. (2009). Lost in translation? Item validity in bilingual political surveys. *Journal of Politics*, 71(4), 1530–1548.
- Pérez, E. O. (2011). The origins and implications of language effects in multilingual surveys: A MIMIC approach with application to Latino political attitudes. *Political Analysis*, 19, 434–454.
- Peytcheva, E. (2008). *Language of administration as a source of measurement error: Implications for surveys of immigrants and cross-cultural survey-research*. University of Michigan.
- Peytcheva, E. (2019). Can the language of a survey interview influence respondent answers? In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)* (pp. 325–337). John Wiley & Sons.
- Peytcheva, E. (2020). the effect of language of survey administration on the response formation process. In M. Sha & T. Gabel (Eds.), *The essential role of language in survey research*. RTI Press.
- Raudenská, P. (2023). Single-item measures of happiness and life satisfaction: The issue of cross-country invariance of popular general well-being measures. *Humanities and Social Sciences Communications*, 10(1), 1–18. <https://doi.org/10.1057/s41599-023-02299->
- Rothstein, B., & Holmberg, S. (2015). Good societies need good leaders on a leash. In C. Dahlström & L. Wängnerud (Eds.), *Elites, institutions and the quality of government*. Palgrave Macmillan.
- Rothstein, B. (2010). *Happiness and the Welfare State*. Social Research: An International Quarterly, 77(2), 441–468.
- Sapir, E. (Ed.). (1929). *The status of linguistics as a science. Culture, knowledge and personality*. University of California Press.
- Schwarz, N. (2003). Culture-sensitive context effects: A challenge for cross-cultural surveys. In J. Harkness, F. J. R. Van de Vijver, & P. Mohler (Eds.), *Cross-cultural survey methods*. John Wiley and Sons.
- Sears, L. E., Agrawal, S., Sidney, J. A., Castle, P. H., Rula, E. Y., Coberley, C. R., Witters, D., Pope, J. E., & Harter, J. K. (2014). The well-being 5: Development and validation of a diagnostic instrument to improve population well-being. *Population Health Management*, 17(6), 357–365.
- Stevenson, B. & Wolfers, J. (2008). Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox, Brookings Papers on Economic Activity, Spring 2008. [https://www.brookings.edu/wpcontent/uploads/2008/03/2008a\\_bpea\\_stevenson.pdf](https://www.brookings.edu/wpcontent/uploads/2008/03/2008a_bpea_stevenson.pdf)
- Suh, E. M., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgements across culture: Emotions versus norms. *Journal of Personality and Social Psychology*, 74(2), 482–493.
- Veenhoven, R. (2008). Comparability of happiness across nations. *School of Sociology and Social Work Journal*, 104, 211–234.
- Veenhoven, R. (2009). The international scale interval study: improving the comparability of responses to survey questions about happiness. In V. Møller & D. Huscicka (Eds.), *Quality of life and the millennium challenge: Advances in quality-of-life studies, theory and research*. Springer.
- Villar, A. (2009). Agreement answer scale design for multilingual surveys: Effects of translation-related changes in verbal labels on response styles and response distributions. In *Survey Research and Methodology program (SRAM)—Dissertations & Theses. Paper 3*. University of Nebraska-Lincoln.
- Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Wiley.
- Wierzbicka, A. (2004). Happiness in cross-linguistic and cross-cultural perspective. *Daedalus, on Happiness*, 133(2), 34–42.
- Wierzbicka, A. (2011). What's wrong with happiness studies? The cultural semantics of happiness Bonheur, Glück and sčas'te. In J. Slovo (Ed.), *World and language*. Jazyki slavjansko kultura.
- Youssef, J., & Diab, S. (2021). Does quality of governance contribute to the heterogeneity in happiness levels across MENA countries? *Journal of Business and Socio-Economic Development*, 1(1), 87–101. <https://doi.org/10.1108/JBSED-03-2021-0027>
- Zwolinsky, J. (2019). Happiness around the world. In K. D. Keith (Ed.), *Cross-cultural psychology: Contemporary themes and perspectives*. John Wiley & Sons Ltd. <https://doi.org/10.1002/9781119519348>