



Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

The econometrics of happiness: Are we underestimating the returns to education and income?

C.P. Barrington-Leigh

Department of Equity, Ethics, and Policy; and Bieler School of Environment, McGill University, 2001 McGill College, Montreal, H3A 1G1, Quebec, Canada

ARTICLE INFO

Keywords:

Happiness
Subjective wellbeing
Life satisfaction
Welfare
Income
Education
Modeling

ABSTRACT

This paper describes a fundamental and empirically conspicuous problem inherent to surveys of human feelings and opinions in which subjective responses are elicited on numerical scales. The paper also proposes a solution. The problem is a tendency by some individuals — particularly those with low levels of education — to simplify the response scale by considering only a subset of possible responses such as the lowest, middle, and highest. In principle, this “focal value rounding” (FVR) behavior renders invalid even the weak ordinality assumption often used in analysis of such data. With “happiness” or life satisfaction data as an example, descriptive methods and a multinomial logit model both show that the effect is large and that education and, to a lesser extent, income level are predictors of FVR behavior. A model simultaneously accounting for the underlying wellbeing and for the degree of FVR is able to estimate the latent subjective wellbeing, i.e. the counterfactual full-scale responses for all respondents, the biases associated with traditional estimates, and the fraction of respondents who exhibit FVR. Addressing this problem helps to resolve a longstanding puzzle in the life satisfaction literature, namely that the returns to education, after adjusting for income, appear to be small or negative. Due to the same econometric problem, the marginal utility of income in a subjective wellbeing sense has been consistently underestimated.

1. Introduction

Now firmly entrenched in the economics literature, in national statistical agency data collection, and in the dialogue about progress and wellbeing, survey-based subjective evaluations of life¹ are the basis for estimating welfare benefits and costs of everything from inflation and unemployment, to air pollution and being married (e.g., Blanchflower et al., 2014; Levinson, 2012; Stutzer and Frey, 2006). Estimates of the psychological benefit of increased income, using this approach, are five decades old, and those evaluating the net individual return of additional education have been carried out for at least three decades. In terms of optimally allocating human resources, not much could be more central than knowing the marginal utility of income and of education.

1.1. Responding to life evaluation questions

However, the coherence and value of subjective evaluations of life rely on a series of considerable cognitive tasks to be performed in

short order by the respondent. When asked,² “Overall, how satisfied are you with life as a whole these days, measured on a scale of 0 to 10?” a respondent must in some sense (i) conceive of the domains, expectations, aspirations or other criteria salient to her sense of experienced life quality or satisfaction; (ii) assemble evidence pertaining to each ideal, such as recent affective (emotional) states, significant events, and objective outcomes; (iii) appropriately weight and aggregate this evidence according to its importance to overall life quality, and (iv) project the result onto the discrete numerical scale specified in the question.

This is without doubt a tall order, and any embrace of **subjective wellbeing (SWB)** data, and especially the headline measure of **life satisfaction (LS)**, rests on their remarkable reproducibility and apparent cardinal comparability, possibly along with the principle that any objective indicator of experienced wellbeing must ultimately be accountable to a subjective one. While various studies have sought, with limited success, to find differences in interpretation of the LS question or norms of expression across cultures and languages (Helliwell

E-mail address: Chris.Barrington-Leigh@McGill.ca.

URL: <https://alum.mit.edu/www/cpbl/address>.

¹ The life satisfaction question and close cousins such as the Cantril Ladder question are posed in numerous national and international social surveys, both cross-sectional and panel. The U.K. Treasury’s Green Book includes instructions for how to use compensating differentials, estimated from life satisfaction, to carry out cost/benefit calculations for central government (UK Treasury, 2021; MacLennan et al., 2021).

² Typically, cognitive evaluations of life consist of a single, subjective, quantitative question like this one, and responses are used directly as a cardinal or ordinal proxy for welfare, i.e., “experienced” utility (Easterlin, 1974).

<https://doi.org/10.1016/j.jpubeco.2023.105052>

Received 16 August 2022; Received in revised form 23 November 2023; Accepted 18 December 2023

Available online 18 January 2024

0047-2727/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

et al., 2010; Exton et al., 2015; Lau et al., 2005; Clark et al., 2005), an important fact is that, uniformly across cultures, responding to the question is cognitively demanding. This paper focuses specifically on the consequences of an apparent heterogeneity across respondents in their ease with the final, quantitative step in the process outlined above.

The crux is that people with less facility with numbers may simplify the numerical response scale for themselves. In particular, the evidence below shows that some respondents restrict the set of numerical options under consideration to a three-point scale consisting of the bottom, middle, and top options, rather than the full set offered. This can be expected to introduce complex biases in mean life satisfaction and in estimated marginal effects on life satisfaction, in particular with respect to education and other correlates of numerical literacy itself.

I present evidence of the prevalence and quantitative significance of this problem, with implications for the interpretation and analysis of all numerical, subjective response scales. The language and empirical examples all focus on the case of single-item SWB questions, mostly LS (Cheung and Lucas, 2014), which underlie the field of the “economics of happiness”. While most empirical studies make use of a cardinal interpretation of the response scale in the life satisfaction question, and at least an ordinality assumption is universal,³ the “**focal value rounding**” (FVR) behavior, described above, introduces a conspicuous violation of the ordinality of response options. Because a number of governments are gearing up to carry out cost/benefit analyses using regressions of LS data for budgeting and program evaluation (Frijters et al., 2020; Frijters and Krekel, 2021; Happiness Research Institute, 2020; Grimes, 2021; Department of Finance, 2021; UK Treasury, 2021), proper econometric accounting for FVR may have practical importance.

In order to estimate the size of systematic biases associated with widely used methods of inference based on SWB reports, I present a model which accounts for the FVR phenomenon and which shows why biases on estimates can be large or small and positive or negative. The model also quantifies the fraction of respondents in a sample who have chosen an alternate, simplified response scale, a value I call the **Focal Value Rounding Index**, or **FVRI**.

The rest of this paper proceeds as follows. The remainder of the Introduction reviews some stylized facts related to education and well-being in the happiness literature, and mentions some points of history in the development of SWB survey questions like LS. Next, Section 2 will convince the reader that there is a measurement problem with quantitative, subjective scales like LS that is conspicuous, ubiquitous, and strongly correlated with educational attainment and that it has a natural explanation supported by the behavioral evidence. Then Section 3 presents the formal model in which a mixture of high- and low-numeracy respondents treat the response scale differently. Section 4 validates the estimation and identification approach using synthetic data and explores the complexity of biases that can result from FVR. Section 5 presents the main empirical estimates of the relationship between education, income, and wellbeing, using a large social survey from Canada. Section 6 reexamines three previously published studies, along with a ranking of U.S. states, as applications to investigate the extent of bias in existing published literature as well as in popular happiness rankings. In these empirical applications, previously anomalous but reproducible findings include evidence that a disadvantaged population reports high life satisfaction, and that the return to extra years of education after primary school are negative, especially when conditioned on income. These surprising findings are overturned when taking into account focal response behavior. A summary and a perspective on future directions are in Section 7.

³ A common finding is that models assuming cardinality give similar results to those assuming only ordinality (Ferrer-i-Carbonell and Frijters, 2004).

1.2. Effects of education and income on subjective wellbeing

Education may be expected to confer welfare benefits not just through higher income but also through better health behaviors and enhanced social capital of various forms with intrinsic benefit (e.g., Helliwell and Putnam, 2007; Powdthavee et al., 2015), as well as through some kind of psychological capital which captures intrinsic benefits of learning or knowledge, or which complements other consumption (for instance, possibly literature, fine art, or the night sky). However, among the more surprising stylized facts in the economics of happiness is that formal education does not help much to explain LS once income⁴ is accounted for (e.g., Layard, 2011; Frijters and Krekel, 2021).⁵

Similarly, although the literature on the importance of income and income growth on LS is enormous and involves a large potential role of consumption externalities (Barrington-Leigh, 2014), one may summarize the findings by saying that income has been found to be a weak predictor of LS in comparison to other, less market-mediated parts of life (e.g., Blanchflower and Oswald, 2004; Hamilton et al., 2016; Layard, 2011; Frijters and Krekel, 2021).⁶

This paper does not aim to explore all the reasons for this well-established evidence about quality of life from subjective response data. Instead, it characterizes a measurement error in which those with lower education and, as a proxy, those with lower income, may be more likely to under-utilize the LS response options in such a way that tends to bias their reported life satisfaction. In general, resulting biases on marginal effects could exist in either direction, but as described below they are more likely to be downward, meaning that they may go some way to explaining the education anomaly and to revise upward, if modestly, the estimated importance of income for supporting SWB.

1.3. Evolution of precision in subjective, quantitative reports

The history of survey questions on subjective assessments mirrors in part technological norms. Early innovators in monitoring SWB in social and household surveys tended to use a four point or five point scale, typically with Likert-style verbal response options. In such questions, the numbers were not meant as cues for the respondent. In some populations, most respondents chose one of the top two options, limiting the variation, or precision. As limitations of paper survey media have been erased by the adoption of computer aided interviews, the resolution of these subjective scales has expanded. However, with more than five or seven response options, verbal cues are typically not provided except for the highest and lowest response options. Responses instead become numerical. For instance, after many years of asking LS questions with a variety of scales, Statistics Canada settled over a decade ago on a particular wording with an 11 point scale.⁷

⁴ Studies routinely control for current income rather than wealth when modeling LS. Current income may be an especially poor proxy for lifetime income in this context because choosing to pursue extra education entails a trade-off between short-term income and future income. This leaves educational attainment as a positive proxy for unmeasured future income expectations, making the low coefficients measured on education even more surprising.

⁵ This generalization hides considerable variation in the literature. Since the various channels and directions of influence are not easily identified, and the relationship may not even be monotonic (Stutzer, 2004), estimates vary from slightly positive to substantially negative overall effects of having extra education.

⁶ That is, the large compensating differentials found for having positive social relationships (trust, engagement, meaningful work, friendships, intimate relationships, etc; see for example Powdthavee, 2008; Helliwell and Barrington-Leigh, 2011; Helliwell and Putnam, 2004) reflect a small denominator, i.e. the value of income for increasing life satisfaction.

⁷ However, Conti and Pudney (2011) describe an evolution in the opposite direction, away from unlabeled response options, in 1992 in the British Household Panel Survey.

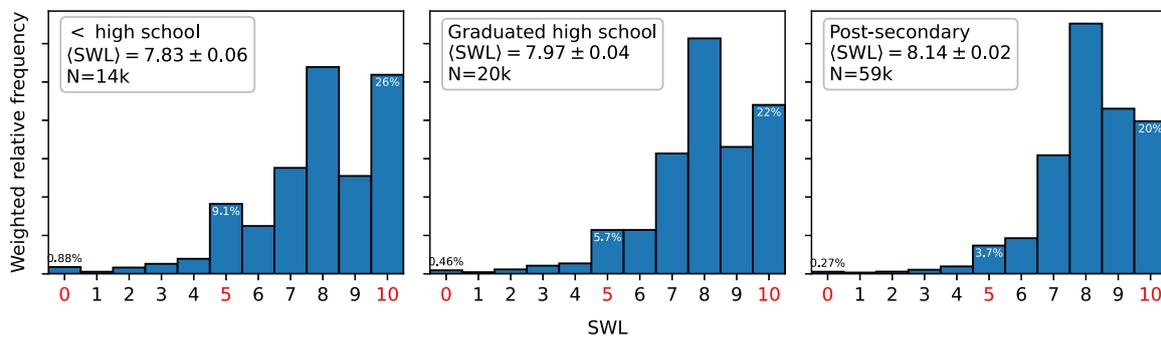


Fig. 1. CCHS: LS responses versus education for those at least 25 years old. Education is coded into three categories. Response fractions within each education category are shown for focal value responses. Use of focal values decreases with increasing education. Mean LS in each group is shown with 95% confidence range. Histograms and means are population estimates, using sampling weights provided by Statistics Canada.

The OECD (2013) has also developed recommendations for standardizing the way such questions are asked by all national statistical agencies. The *de facto* standard for LS now is an 11-point scaling, from 0 to 10, with the lower extreme meaning, for example, “not at all satisfied”, the upper signifying “completely satisfied”, and the interpretation of the remaining values left up to the respondent.

An older literature sought to determine the optimal number of response options in survey questions with verbal cues for each option. For instance, it may be that in an oral interview, i.e., with no visual cues, four or five responses are the maximum that can be handled without confusion or overload (Bradburn et al., 2004).

When the scale is explicitly numeric, as with modern LS measures, there also arises a trade-off between the cognitive load imposed by a scale and the precision it allows. From the respondent’s point of view, this trade-off is between the opportunity for self-expression and the cost of cognitive processing. The survey designer wishes to allow for precise responses in order to capture variability among respondents and over time, while not demanding too much. Overburdening would result, at best, in the respondent not fully optimizing her answer or not properly interpreting or using the given range of responses (OECD, 2013). Various studies on this balance have tended to favor 11-point quantitative scales over coarser option sets (e.g. 7-point scales) as well as over nearly continuous options (Alwin, 1997; Kroh et al., 2006; Saris et al., 1998; OECD, 2013; Weng, 2004).⁸

2. Descriptive evidence

A small number of studies have remarked in some way on the use of focal values, but without a full account or explanation.⁹ Dolan et al. (2011) mention that LS ratings in one study are positively associated with life circumstances as one would expect, except at the top of the scale, where “those rating their life satisfaction as ‘ten out of ten’ are older, have less income and less education than those whose life satisfaction is nine out of ten”. They speculate a reason unrelated to cognitive limitations for this observation but declare that “This issue warrants further research”. Conti and Pudney (2011) describe focal value behavior as a response to the existence of verbal cues, present

⁸ Interestingly, government surveys in the USA have tended to stick with 3 or 4-point scales for SWB questions.

⁹ There is also a psychometrics literature which refers to the tendency towards top and bottom value responses as “extreme response style” and tendency towards the central value as “moderate response style” (Hamamura et al., 2008; Khorramdel et al., 2019). That literature, known in psychology as “item response theory”, is motivated by an interest in personality type, as classified by responses to a battery of Likert questions with all-verbal response options. These studies have not considered cognitive ability as an explanatory factor. Giustinelli et al. (2020) also study rounding of reported quantitative beliefs which relies on observing multiple questions for each respondent.

on only three out of seven response options. Landua (1992) analyses response transition probabilities in the German Socio-Economic Panel, and Frick et al. (2006) confirm his report that respondents have a tendency to move away from the endpoints over time. In fact, this could be driven largely by the FVR behavior diminishing as panel participants, especially those with low numeracy, gain familiarity and comfort with the scale.

2.1. Educational attainment

Simply inspecting their LS distributions, stratified by education level, might have led these authors to the hypothesis developed in this paper. For illustrative purposes I appeal to one cycle from the Canadian Community Health Survey (CCHS), a large annual cross-section which includes an 11-point life satisfaction question as well as educational attainment.¹⁰ Conditioning SWB responses on educational attainment reveals a striking feature (Fig. 1). The relative frequencies of each focal value (0, 5, and 10) decrease with increasing education level. While the lowest education category shows four peaks, the distribution of responses in the highest education category features what would be a unimodal distribution around SWL=8, except for a slight enhancement at SWL=0. In addition to Fig. 1, several other lines of evidence support the interpretation that scale simplification is a specific response to cognitive challenge, a model to be formalized in Section 3.

2.2. Numeracy

Educational attainment is a widely-available characteristic in social survey data, but may capture attributes which are relevant to latent wellbeing or to reporting functions, but which are different from cognitive ability with numbers. To support the numeracy interpretation of FVR, Appendix Figure F.1 makes use of the 2018 Programme for International Student Assessment (PISA) survey of 15 year old enrolled school students in 72 countries. This survey includes a measure of mathematical ability, along with 11-point life satisfaction. Separating respondents according to an overall math score,¹¹ the same feature as in Fig. 1 is observed: the relative frequency of each focal value response decreases with math proficiency, as does the mean LS response. This is true of the global distribution, as well as for individual countries such as the USA (see Appendix Figure F.1).

¹⁰ In a repeated cross-section, most respondents are facing the LS question for the first time. In the CCHS, education is recorded in four categories: less than high school graduate, high school graduate, some post-secondary training, and completed post-secondary training, but relatively few respondents report the third category, so I combine the top two. In the 2017–2018 wave, out of a total of 113 289 respondents, 97 604 reported their age as at least 25 years, and 93 043 of those answered the LS, educational attainment, and income questions.

¹¹ This math score is a “plausible value” of the individual’s underlying latent ability, appropriate for modeling relationships such as this one; see Wu (2005).

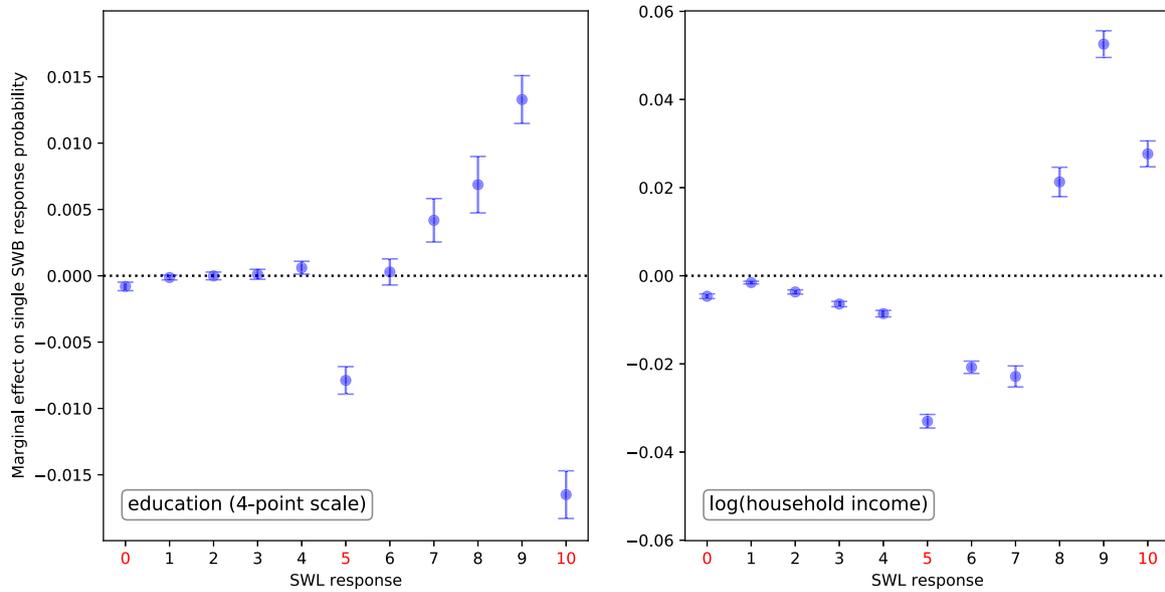


Fig. 2. Multinomial logit estimate of individuals' probability of giving each possible response to the life satisfaction question (CCHS data). For educational attainment (quantified on a 1–4 scale) marginal effects show a monotonic pattern with increasing response value, except for the remarkable outliers at 0, 5, and 10. These indicate that education and, simultaneously but to a lesser degree, income are significant predictors of the tendency to use a simplified response scale. Error bars denote 95% confidence intervals.

2.3. Difficulty responding

Another indication that the apparent tendency to simplify the response scale has to do with the difficulty of answering the question, as it is posed, comes from noticing that respondents with less education are more likely to refuse to answer the LS question at all. Appendix Table F.1 shows that response rates to the LS question, although close to 100%, are strictly increasing with educational attainment.

2.4. Unordered choice model

The existence of FVR behavior implies that SWB response scales cannot safely be assumed to be ordinal. For example, those with lower education may, all else equal, experience lower life satisfaction but be systematically inclined to report a higher value due to rounding up from a 3 or 4 to 5, or from 8 or 9 to 10. It is possible, therefore, that on average those reporting 9 could be happier than those reporting 10.

Traditional methods used in econometric inference from LS — such as OLS, ordered logit, ordered probit, and related time series and instrumented analogues — leverage strong assumptions about the symmetry of effects of explanatory variables on each step of the response scale, as well as assuming cardinality or at least ordinality among response values. Those models are therefore not flexible enough to account for the heterogeneous influence of predictors like education on focal and non-focal response values.¹²

An alternative approach is to relax the ordinality assumption for response options, and model the probability of each response independently, subject only to the constraint that the probabilities add up to one. The multinomial (polytomous) logit model¹³ does this.

¹² Note that, unlike OLS, ordered logit and ordered probit naturally account for multi-peaked distributions such as is shown in Fig. 1. That is due to the flexibility given by the cut points in those models, which can squeeze together or stretch apart in order to decrease or increase (respectively) an option's predicted response probability.

¹³ The multinomial logit model, in its latent variable formulation, consists of a system of equations generating scores $Y_{i,j}^*$ for each individual i and response option $j \in \{0 \dots 10\}$ as $Y_{i,j}^* = \beta_j \cdot x_i + \varepsilon_j$ where $\varepsilon_j \sim EV_1(0,1)$, i.e., the error terms have standard type-1 extreme value distributions. Then observation probabilities are given by $P(Y_i = s_j) = e^{\beta_j \cdot x_i} / \sum_{k \neq j} e^{\beta_k \cdot x_i}$ with one necessary

Fig. 2 shows marginal effects of education and income on response probabilities of each of the 11 points in the LS scale, from a multinomial logit model using education, logarithmic income, age, and age² as predictors for the sample shown in Fig. 1. Under an ordinality assumption, one might expect marginal effects to rise monotonically with response value, since a better circumstance like education or income should lead to an increase in the relative probability of response $s + 1$ as compared with response s . Indeed, other than the focal response values 0, 5, and 10, the marginal effect of one step higher educational attainment (for instance, graduating from high school) is weakly increasing in reported LS. By contrast, the effects on the focal value responses are, with high statistical significance, negative¹⁴ outliers far below what would be expected based on the pattern of adjacent values. They show that more education significantly *reduces* the probabilities of each focal value response. Multinomial logit estimations provide a diagnostic tool for detecting predictors of focal value behavior. However, the effect sizes are hard to interpret because they are averages over the entire sample. For instance, the education coefficient for LS=10 is an average effect over *high* types, for whom higher education increases the chance of reporting 10, and *low* types, for whom higher education decreases that chance. In order to separate those effects, a more structured mixture model approach, described below, is required.

2.5. Precision and self-expression

As a final piece of empirical motivation for the modeling approach developed below, I note that when excess precision is offered in an SWB scale, respondents appear to make a costly effort to choose round numbers. Specifically, Appendix Figure F.2 shows the distribution of responses from a computer-based SWB survey question framed on a 0–10 scale but with an available resolution of 0.1. There are clearly favored responses at every integer and half-integer value. The response

normalization like $\beta_0 = 0$. This model is clearly also misspecified for LS data, since the formal *independence of irrelevant alternatives* (IIA) assumption is violated by the numbered options of an LS question. The IIA requirement is frequently overlooked in applications of the multinomial logit.

¹⁴ For each explanatory variable, the sum of all marginal effects on probabilities is zero by construction.

interface was a graphical slider which gave no preference for any particular values. Thus, the prevalence of rounded values indicates that *extra effort* in the form of fine manual control was exerted in order to leave the slider precisely on a half- or whole-integer value. This can be interpreted as evidence of effort to faithfully communicate a mental result, motivated by the drive for self-expression (Alwin, 1997; OECD, 2013).¹⁵

3. Cognitive mixture model

Motivated by the evidence above, the enhanced use of focal values can be interpreted as an indication that respondents have simplified their cognitive task by coarsening the numerical scale. Because FVR behavior is inversely associated with education and math skills, I focus on “numeracy” as one major influence on scale choice. The two-type mixture model below is based on the assumption that the cognitive processes of respondents differ only in the execution of step (iv) described in the second paragraph of Section 1. That is, an internal representation of overall wellbeing exists in a similar way across the two groups, who subsequently project that assessment onto either the full scale (*high numeracy* respondents) or a subset consisting of the bottom, central, and top values (*low numeracy* respondents).

For each of the two types, latent wellbeing is mapped onto a discrete response scale as in a standard, i.e. canonical, ordered logit model. That is, given a continuous, latent subjective assessment S^* modeled in terms of individual characteristics x as $S^* = x'\beta_S + \varepsilon$, the cumulative probability of discrete responses k is given by:

$$P(s | x) = P(S^* > \alpha_k | x) = P(x'\beta_S + \varepsilon > \alpha_k) \\ = 1 - \Phi(\alpha_k - x'\beta_S) \quad (1)$$

where α_k are a sequence of threshold values α_k^H separating the full set of observed responses $\{0, 1, \dots, 10\}$, or α_k^L for the focal subset $\{0, 5, 10\}$, and $\Phi(\cdot)$ is the cumulative distribution function of the unexplained portion ε of S^* . Use of the logistic distribution for $\Phi(\cdot)$ makes this an ordered logit model.

The *high* and *low* alternative ordered logit outcomes are combined using a simple dichotomous logit model. If z is a vector of individual characteristics, possibly overlapping with x , which serve as a measure of numeracy, then

$$P(\text{high} | z) = \Phi(z'\beta_N) \quad (2)$$

There is no explicit consideration of costs and benefits to the respondent.¹⁶

Together, Eqs. (1) and (2) form a mixture model. The probability of observing response k is

$$P(k | x, z) = P(\text{high} | z) P(k | x, \text{high}) \\ + [1 - P(\text{high} | z)] P(k | x, \text{low}) \quad (3)$$

¹⁵ It also suggests that respondents have introspective knowledge of their degree of precision in answering a numerical SWB question — knowledge which is not typically elicited in surveys.

¹⁶ Conceptually, the individual benefits of using the full scale are self-expression and performing one’s best at fulfilling the purpose of the survey; the costs are those of the cognitive calculation. For more on the tradeoff between self-expressive capacity and cognitive capacity in the design of response scales, see Alwin 1997, Kroh et al. 2006, Saris et al. 1998, OECD 2013. There is no evidence that respondents’ value of time, which might for instance vary with income, is also a major factor in scale choice in the face of the inclination for self-expression. The value of time likely affects the choice to participate in a survey, but once asked the LS question, respondents answer it quickly, i.e., in a few seconds. The indication in evidence shown above and below is that higher income predicts higher resolution in responses, which seems contradictory to a hypothesis of behavior driven by the market value of respondents’ time.

The model is similar to the ordinal-outcome “finite mixture model” of Boes and Winkelmann (2006) except that here the mixing probability is dependent on individual characteristics (see also Everitt and Merette, 1990; Everitt, 1988; Uebersax, 1999). A more detailed account of the model is presented in Appendix A.

3.1. Identification

Are the parameters in this model point-identified in principle?¹⁷ Identification is a challenge because the same predictors may be used to predict the latent numeracy variable and to predict the latent wellbeing variable. As a result, one might fear that more than one set of parameters could equally well explain observations for a given sample. Excluding the columns of z from x in (3) would overcome this problem. However, for an all-encompassing subjective outcome such as latent wellbeing, it is safer to assume that everything could be a determinant. More specifically, a particular interest motivating this study is to assess the bias on estimates of the wellbeing effect of education, and education is also the primary available predictor of numeracy.

With stronger assumptions, an alternative strategy to the mixture model may be able to identify parameters for latent SWB by avoiding FVR altogether. One approach would be through thin set identification, if respondents with some level of education were known never to exhibit FVR. One standard problem with this kind of identification is that it relies on an assumption of a uniform effect of a covariate across its support, as well as the absence of interaction effects with other covariates. By contrast, the mixture model approach of (3), which leverages the entire sample, has the advantage of generalizability to explicitly estimate interaction terms or other functional forms to allow for non-uniform effects.

Another approach would be through selection on the dependent variable; that is, by restricting the sample to the subset of high types who did not respond with a focal value. Because no “5”s are observed in this group, it would consist of two subsamples: those with observed $s \in \{1, 2, 3, 4\}$ and those with $s \in \{6, 7, 8, 9\}$. In fact, if the symmetries required for this approach to be unbiased were believed, then one could likely estimate coefficients for latent wellbeing using binary models like logit and sample subsets of respondents who answered one of only two consecutive, non-focal response options.

Returning to (3), within each of the two ordered logit formulations nested in the model, identification of the set of parameters (with no constant term) is standard. This still leaves us with incomplete identification, in general, of the parameters on variables common to x and z . One can imagine extreme distributions of SWB, for instance all near 10, in which FVR only acts to convert 9s to 10 s. In this case, discriminating between the effect of common variables on latent wellbeing or FVR would not be possible, especially if the sign of coefficients is not constrained. However, more typically, with a broader SWB distribution, FVR will be distinguishable from effects on latent wellbeing. That is, successful identification rests on having sufficient independent, explainable variance in latent SWB across low types in order that there is also variation in their observed response. For instance, if the latent wellbeing of low types is sufficiently spread out that they sometimes round down and sometimes round up, then the influence of education on SWB, controlling for other influences, is separately identified from the influence of education on the reporting function, i.e., on the likelihood of being a low type. Put differently, identification

¹⁷ Point-identification, typically referred to simply as “identification”, is also called frequentist identification and is a frequentist concept. In Bayesian estimation, as is used in the empirics to follow, parameters are assumed to have distributions, not point values. Using a Bayesian estimation method with a broad enough prior, alternate sets of values which account for the data are simply reflected in multimodal (or suitably broad) estimates of the parameters (Lewbel, 2019).

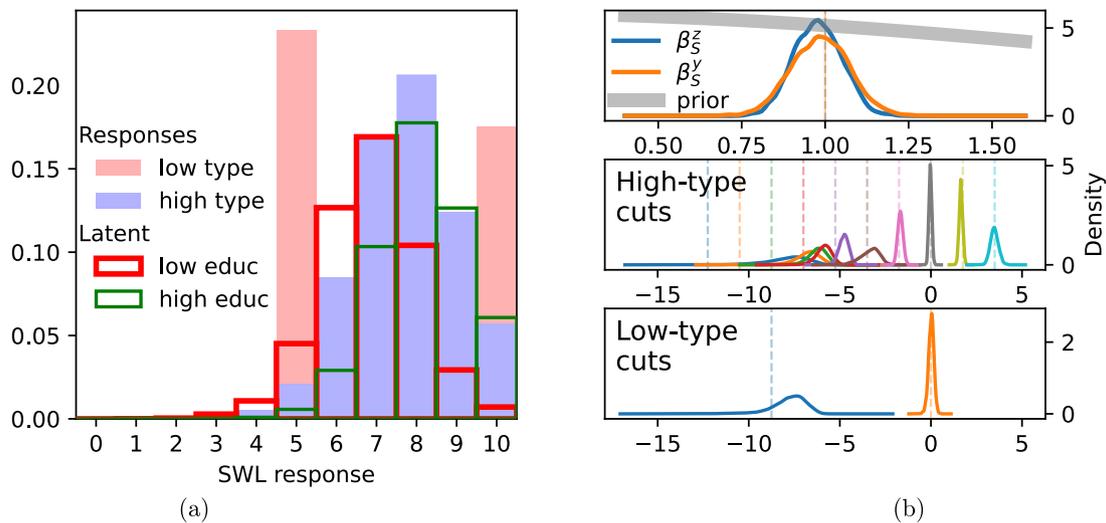


Fig. 3. Example of synthetic data and validation. Panel (a) shows simulated and latent responses for one set of synthetic parameters. Overall in this example, 33% of respondents are low type. They reported an average SWB of 6.8, rather than their true average of 7.0, while high types reported an average of 7.7. Panel (b) shows that the FVR mixture model correctly recovers synthetic coefficients β_S and cut points α^H , and α^L . Vertical dashed lines show the true (data generating process) values. Most cut points are precisely estimated, but the lowest ones are poorly constrained because there are few low responses for this particular set of synthetic parameters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

comes from the response of the observed distribution to changes in numeracy, driven by some variable, being different from the response of the observed distribution to changes in latent wellbeing, driven by the same variable. This is assured when there is nontrivial variation in the latent wellbeing of low types. This conceptual argument is best corroborated quantitatively through simulation, which demonstrates, in Section 4, that β_S and β_N are simultaneously recovered when estimating Eq. (3).

3.2. Focal Value Rounding Index

As shown below, net biases on some estimated moments and model coefficients may be zero due to offsetting effects, even when FVR behavior is prominent. Therefore, to express straightforwardly the magnitude of the numeracy problem in a sample of respondents, another estimated value is helpful. This is the Focal Value Rounding Index (FVR), which is an estimate of the fraction of the population who restrict their answer to a set of focal values — i.e., the estimated fraction of low types. This value is well identified whenever β_N is.

3.3. Counterfactual SWL distribution

The mixture model provides a posterior estimate of the latent SWB distribution, i.e., that which would have been reported had respondents all used the full scale. This represents a “correction” to the reported distribution of SWB. This is a distribution of predicted, counterfactual, discrete responses on the 0–10 scale, not an estimate of the latent variable S^* . The next section demonstrates through simulation that the model successfully recovers (identifies) this counterfactual distribution, along with the FVR, means, and coefficients.

4. Model validation

This section, supplemented by several appendices, reports on the use of simulated data to validate the computational approach¹⁸ and the

¹⁸ Estimation of the mixture model was carried out using the no-U-turn sampler (NUTS) variant of a Hamiltonian Monte Carlo (HMC) algorithm, which is in turn a Markov Chain Monte Carlo (MCMC) method (Stan Development Team, 2018; Riddell et al., 2020; Carpenter et al., 2017) and handles the

model’s ability to identify simultaneous influences of a predictor, like education, on FVR and on latent SWB. A large battery of simulations demonstrates the complexity and scope of possible biases, due to FVR, in conventional estimates of SWB means and of marginal effects.¹⁹

4.1. Synthetic data validation results

Simulated SWB data are generated by a model in which a scalar z partly determines numeracy through (2) while z and a second scalar, y , partly determine the latent wellbeing S^* (thus $x \equiv \begin{bmatrix} z \\ y \end{bmatrix}$ in (1)). A non-zero correlation, parameterized by χ , may exist between z and y . Conceptually, and for comparison with the empirical estimates to follow, z is meant to represent education and y represents other variables, such as income, which are not direct measures of numeracy (do not cause FVR). In order to reveal the possible scope of biases for plausible distributions of SWB, a number of parameters of the synthetic data generation process were varied systematically. These include χ , β_N , and two parameters determining the *scale* and *offset* of the cut points.²⁰

Fig. 3 shows one example of a simulated distribution of SWB. In (a), shaded bars represent simulated responses on a 0 to 10 scale. Unlike in real data, we are able to identify which respondents (among those giving a 0, 5, or 10) used FVR. This portion of responses, labeled “low type”, are shaded pink. Also because the data are synthetic, we are able to construct the latent (“true”) wellbeing levels and thus the counterfactual 0–10 responses which would have been given if everyone reported

non-concave objective well. Appendix B provides more detail on estimation, including analytic derivations of the gradient and Hessian for a log likelihood approach.

¹⁹ To reiterate, this bias is, conceptually, the difference between naively estimated values and those which would be obtained in the counterfactual case that all respondents had used the full numeric response scale, i.e. were of “high numeracy” type. This counterfactual can be perfectly calculated using synthetic data, but is of course unobservable in traditional empirical data.

²⁰ See Appendix C for details of the parameters used in the synthetic data generating function, Appendix D for some detail from estimates using the simulated data, showing an example of the complicated dependence of biases on attributes of the distribution, and Appendix E for propositions on the maximum possible scope of these biases.

without FVR. This counterfactual distribution, including both low and high types, is shown split into two groups based on education level. Although the true wellbeing distribution of this sample is centered around 7.5, equidistant from the focal values of 5 and 10, there is a net negative bias of -0.08 in mean reported SWB. This is because the distribution of the lower educated component is generally closer to “5” than to “10”. Thus, the amount of rounding up is less than the amount of rounding down.

Simulated biases in regression coefficients are obtained by estimating a traditional ordered logit model on the synthetic data, and comparing those estimates to the true values used in constructing the data, $\beta_S^z = \beta_S^y = 1$. In the case shown in Fig. 3, these biases are also both negative, namely -14% and -23% respectively.²¹

Simulations were carried out for a wide range of parameters, generating cases with both positive and negative biases on mean LS and on β_S^z much larger than in this example. Simulated biases on β_S^y , by contrast, tend to be negative.²² More generally, the bias on mean LS can be as large as ± 2 points (see Appendix Proposition E.1 in Appendix E) and the bias on β_S^z may be even larger (Appendix Proposition E.1). In all cases, the true distribution, fraction of low-types, and effects of z and y on latent wellbeing are identified and correctly estimated by the FVR mixture model. As an example, Fig. 3(b) shows estimated coefficients and cut points for the same case shown in (a).

4.2. Variance of SWL (“happiness inequality”)

Although not a focus of this paper, it is also worth noting that variance of SWB, which has attracted interest as a measure of inequality (Goff et al., 2018; Hasegawa and Ueda, 2011; Stevenson and Wolfers, 2008), also suffers from bias due to FVR, as of course do other moments and other measures of dispersion. For a relatively narrow distribution of LS centered around 5, FVR behavior decreases the variance. For a wider distribution, focal values of 0 and 10 would become prominent, and the variance could be biased upwards instead.

5. Empirical estimates of FVR bias and FVR

With the above evidence of parameter identification from simulated data, the rest of this paper turns to empirical estimates. The distributions of LS for different levels of education, shown in Fig. 1, indicate the significance of focal value rounding behavior in the CCHS sample. Using the mixture model, the role of education in supporting LS can be estimated, despite the strong relationship between education and the focal value bias. Columns (1) and (2) of Table 1 show the results of conventional, or “naive” estimates of the following simple individual-level cross-sectional OLS model for LS,

$$SWL_i = c + \sum_j \beta_S^{h_j} \text{education}_{j,i} + \beta_S^l \log(\text{HH income})_i + \varepsilon_i \quad (4)$$

as well as its ordered logit counterpart. Educational attainment is captured by a set of cumulative dummies, so that β^{h_j} is the impact of having completed education level j or higher.

The naive estimated coefficient on completing secondary education is near-zero or distinctly negative in the two estimates. The ordered logit coefficients predict that completion of high school *reduces* the odds

²¹ The origin of these negative biases is slightly more subtle; see Appendix D and Appendix E for details.

²² See Appendix D for an explanation. In the simulations, y has no extra effect on (information about) numeracy, after taking z into account. In real data, income is likely to contain variance that is informative for FVR but orthogonal to available measures of education. In this case, income in empirical applications will exhibit a blend of the bias features attributed to z and y in these simulations.

of a higher LS by more than 7%, and that even a university education reduces those odds by nearly 4% as compared with someone who has less than a high school education. These values are economically large; using the simultaneously-estimated coefficient on log income, the former effect is estimated to be equivalent to a 13% reduction in income.²³

When constrained to disallow focal value behavior, the mixture model’s estimate, shown in column (3) of Table 1, reproduces the ordered logit values, as it should. However, when the full model is estimated, a significantly positive value (~ 0.06) is found for the LS benefit of completion of secondary school, and an additional 0.17 for those completing post-secondary.

The bias in a conventional estimate of the income coefficient is also large: the mixture model strongly rejects the naive estimated value of ~ 0.53 , in favor of a value of ~ 0.62 . This represents a 17% difference in the most studied value in happiness economics. Combining these coefficients implies that, after controlling for income, the true benefit of college completion, as compared with an otherwise-similar respondent without high school completion, is equivalent to an additional 45% of income. High school completion by itself confers a benefit equivalent to more than 10% of income, after controlling for differences in actual income.

The specification in Table 1 includes both education and income as predictors of FVR. Appendix Table F.2 shows that alternate models with only education in the FVR equation, or with additional controls, give highly consistent results.

Next to Table 1 are visualizations of several sets of distributions, showing the model’s ability to predict observed response patterns while estimating the distribution of “underlying” or “true” SWB.

6. Applications

Hundreds of empirical papers estimating models of life satisfaction and other extended-Likert-like scales could be revisited in light of the significant possibility of biases identified above. Those focusing on effects of socioeconomic status, gender, and age, and those which particularly address populations with low levels of numeracy, especially invite reanalysis. Here I reproduce estimates from three papers to exemplify the important changes that may result from such analysis, and to show that the often-reproduced “paradox” of negative benefits to education may be largely resolved by the cognitive mixture model.

6.1. U.K.: Clark and Oswald (1996)

The first of these papers, with over 1500 citations, is a relatively early contribution in the modern study of relative income concerns but also prominently points out the anomalously low estimated returns to wellbeing from education (Clark and Oswald, 1996). It was also recently cited as one of 11 studies in the major accumulated evidence on the life satisfaction benefits from additional education (Clark et al., 2019, see Annex 3a). In fact, the paper uses data from the British Household Panel Survey (BHPS) prior to its inclusion of LS, so it uses instead responses to 7-point satisfaction with pay and satisfaction with job questions.

Clark and Oswald (1996) did not examine the distributions of these subjective response variables according to formal educational attainment.²⁴ Doing so reveals dramatic FVR behavior which roughly

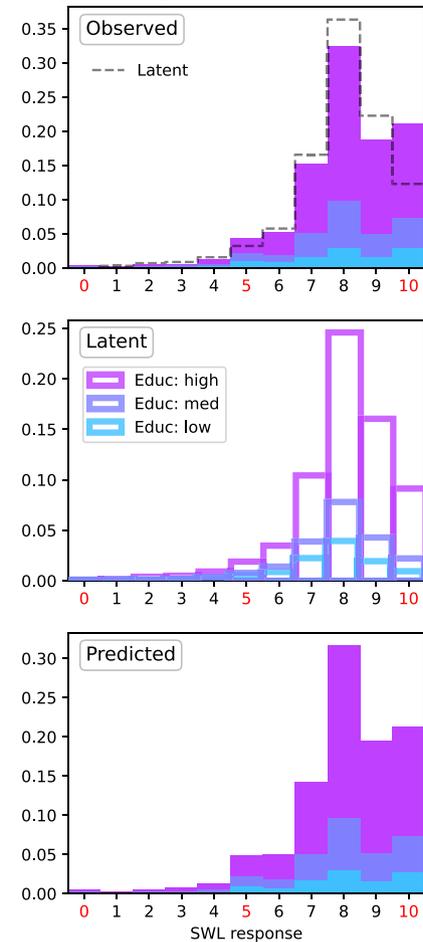
²³ The values in this paragraph are calculated as $e^{-0.075} - 1 = -0.072 \approx -7\%$; $e^{-0.075+0.038} - 1 = -0.036 \approx -4\%$; and $e^{-0.075/0.53} - 1 = -0.13 \approx -13\%$.

²⁴ The description from Clark and Oswald (1996) reads: “Table 5 contains two ordered probits, in each of which three dummies for educational attainment are included as well as a control for income. The dummies are for a college degree, advanced high school (A-level approximately), and intermediate high school (O-level approximately). The omitted category is for no or low qualifications. These four categories are for achieved paper certificates and not merely for years of schooling”.

Table 1

Estimates of life satisfaction in CCHS. The first two columns show conventional, “naive” estimates (as raw coefficients) of a model explaining life satisfaction with just education and household income. Column (3) shows estimates from a degenerate version of the mixture model constrained to exclude FVRbehavior. Column (4) shows the unconstrained mixture model estimate, with significantly higher effects of education and income on life satisfaction. Histograms show response distributions split up (and colored) by education. The top plot is observed values, with the model’s inferred overall latent distribution shown by a dashed line. The second and third show latent (or “corrected”) and predicted responses.

	Conventional		Mixture model	
	OLS (1)	ologit (2)	FVRI→0 (3)	Mixture (4)
Life satisfaction (β_S)				
School: \geq Secondary	-.002 (.019)	-.075* (.023)	-.073 [†] (.021)	.060* (.024)
School: Post-secondary	.069 [†] (.014)	.038 (.015)	.038* (.015)	.17 [†] (.018)
log(HH income)	.55 [†] (.009)	.53 [†] (.011)	.53 [†] (.010)	.62 [†] (.011)
constant	2.0 [†] (.093)			
Numeracy (β_N)				
constant				1.09 [†] (.043)
School: \geq Secondary				.45 [†] (.038)
School: Post-secondary				.53 [†] (.045)
log(HH income)				.22 [†] (.028)
FVRI				.14 [†] (.008)
obs. log likelihood	91796 -177641	91796 -160572	91796 -160597	91796 -159586
Significance:	0.1% [†]	1%*	5%	10% ⁺



diminishes with education (Fig. 4). The distribution of satisfaction with pay is wider and more central (i.e., near “4”) than that of job satisfaction, and features unmistakable evidence of all three focal values (1, 4, and 7) for groups with lower academic certifications. For job satisfaction, the upper focal value is most obvious but all three are evident on inspection. If those with A-levels but no College are excluded, then the group means and the prevalence of each focal value all decrease monotonically with education.

Table 2 shows raw coefficients for model estimates of overall satisfaction with job. The first three models are conventional estimation approaches, including an ordered probit model, which nearly reproduces the published values²⁵ and retained sample size (4730 in all my

estimates) of the main estimate in Clark and Oswald (1996, Table 5).²⁶ In ordered probit, OLS, and ordered logit models, academic attainment is strongly predictive of lower satisfaction after adjusting for log of household income. The implied effect is enormous. As compared with someone with primary education only, an advanced high school graduate (A-levels) is less satisfied with their job by as much as they would be with a 3-fold decrease in wage.²⁷ As already shown in Fig. 4, even the raw mean job satisfaction is decreasing across the first three education groups. Clark and Oswald (1996) speculate that their findings of low satisfaction of the higher educated may be related to a recent recession that particularly hit the middle class in the UK, but also cite several earlier studies which corroborate the negative or negligible benefits from education on job satisfaction.

Equally surprising in these results is the nil effect of income on job satisfaction. The 95% confidence interval for the coefficient of log income in column (3) is -0.10 to + 0.13, with the upper limit

²⁵ The coefficient shown on log income (.016) strongly disagrees with the published value (.50) in Clark and Oswald (1996). Upon contacting the authors, it was determined that a typo in production of the original work resulted in a reporting of 0.50 rather than the estimated 0.05 for this coefficient (personal communication, Andrew Clark, 2021). This error has not been previously reported. Because of the error, the authors did not address the surprisingly low coefficient on log of household income. The set of regional, health, race, industry, and occupation dummies are excluded in Table 3 because the exact definitions from the 1996 work are not available.

²⁶ For easier comparison with their table, the education categories are mutually exclusive, rather than cumulative, as in the CCHS data and the data to follow.

²⁷ The coefficients on log job hours and on A-levels education are nearly identical, implying that, having already adjusted for income, a unit log, or factor ~2.7, increase in hours worked predicts a similar change to job satisfaction as does the educational attainment.

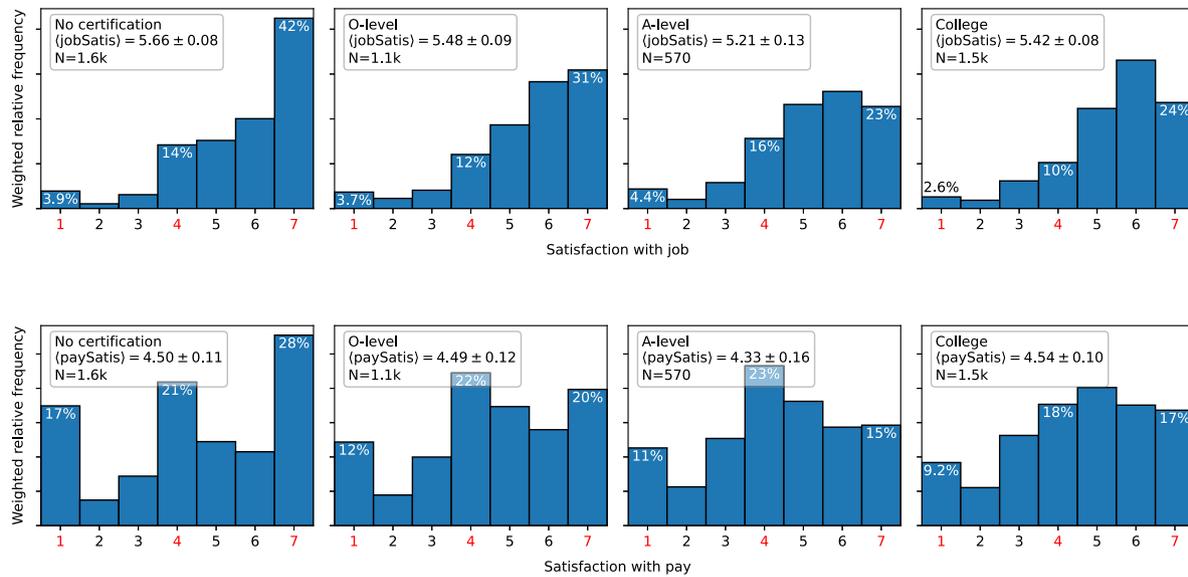


Fig. 4. Distributions of satisfaction with job and with pay for different categories of educational attainment.

Table 2

Estimates of job satisfaction in BHPS. Raw coefficients are shown. Education indicators identify mutually exclusive groups in comparison to those with less than O-levels.

	Conventional			Mixture model	
	oprobit (1)	OLS (2)	ologit (3)	FVRI→0 (4)	Mixture (5)
Job Satisfaction (β_j)					
log(HH income)	.016 (.036)	.074 (.049)	.015 (.060)	.006 (.060)	.14 (.076)
EDUC: College degree	-.15[†] (.046)	-.14 (.061)	-.27[†] (.078)	-.25[†] (.075)	.001 (.11)
EDUC: A-levels (approx)	-.26[†] (.055)	-.31[†] (.076)	-.45[†] (.095)	-.43[†] (.094)	-.24 (.11)
EDUC: O-levels (approx)	-.10 (.046)	-.096 (.060)	-.19 (.078)	-.18* (.075)	.013 (.092)
Log job hours	-.27[†] (.054)	-.35[†] (.071)	-.45[†] (.092)	-.44[†] (.088)	-.46[†] (.099)
age	-.033[†] (.009)	-.037* (.012)	-.055[†] (.015)	-.051[†] (.015)	-.051* (.017)
age ² /1000	.53[†] (.11)	.59[†] (.15)	.88[†] (.19)	.83[†] (.18)	.84[†] (.21)
female	.24[†] (.036)	.35[†] (.050)	.41[†] (.061)	.41[†] (.062)	.45[†] (.069)
constant		6.3[†] (.32)			
Numeracy (β_N)					
constant					.41* (.16)
log(HH income)					.49[†] (.084)
EDUC: College degree					1.30[†] (.28)
EDUC: A-levels (approx)					.86[†] (.21)
EDUC: O-levels (approx)					.69[†] (.14)
FVRI				0[†] (0)	.28[†] (.036)
obs.	4730	4730	4730	4730	4730
log likelihood	-7522	-8573	-7516	-7531	-7466

implying that a doubling of income would increase the odds of a higher satisfaction response by less than 10%.

Column (4) simply shows that the cognitive mixture model reproduces an ordered logit estimate when focal value behavior is turned off, while the key result lies in Column (5). When focal value behavior

is accounted for, the income coefficient increases to a confidently positive value, and the strongly negative coefficients on O-level and College completion are eliminated. Respondents who finished A-levels but stopped there for some reason, i.e., did not complete college, are still predicted to be less satisfied with their jobs, but the effect is half

Table 3
Estimates of satisfaction with pay in BHPS. Description as for Table 2.

	Conventional			Mixture model	
	oprobit (1)	OLS (2)	ologit (3)	FVRI→0 (4)	Mixture (5)
Pay Satisfaction (β_S)					
log(HH income)	.50 [†] (.038)	.92 [†] (.062)	.88 [†] (.065)	.87 [†] (.059)	.96 [†] (.068)
EDUC: College degree	-.17 [†] (.045)	-.26 [†] (.077)	-.31 [†] (.078)	-.30 [†] (.071)	-.26* (.089)
EDUC: A-levels (approx)	-.14* (.053)	-.20 (.097)	-.25* (.090)	-.24* (.091)	-.20 (.10)
EDUC: O-levels (approx)	-.029 (.045)	-.019 (.077)	-.051 (.077)	-.042 (.072)	.005 (.081)
Log job hours	-.82 [†] (.058)	-1.42 [†] (.089)	-1.44 [†] (.10)	-1.43 [†] (.091)	-1.46 [†] (.094)
age	-.043 [†] (.009)	-.072 [†] (.015)	-.077 [†] (.015)	-.071 [†] (.014)	-.074 [†] (.015)
age ² /1000	.62 [†] (.11)	1.03 [†] (.19)	1.10 [†] (.19)	1.03 [†] (.18)	1.06 [†] (.18)
female	.27 [†] (.035)	.48 [†] (.064)	.45 [†] (.059)	.44 [†] (.059)	.45 [†] (.064)
constant		3.8 [†] (.40)			
Numeracy (β_N)					
constant					.22 ⁺ (.17)
log(HH income)					.49 [†] (.075)
EDUC: College degree					1.26 [†] (.23)
EDUC: A-levels (approx)					1.10 [†] (.24)
EDUC: O-levels (approx)					.72 [†] (.15)
FVRI				0 [†] (0)	.31 [†] (.041)
obs.	4730	4730	4730	4730	4730
log likelihood	-8642	-9689	-8635	-8648	-8540

as large as in the naive model. Estimates of other coefficients remain statistically unchanged. Both formal education and reported income prove significant in predicting focal value behavior. The estimated fraction of respondents, overall, who restricted their answers to focal values is 28%. The model also estimates a significant bias in the mean reported job satisfaction, from a latent value of 5.3 which would have obtained had all respondents used the full scale, to the observed value of 5.5. The model estimates that the low-numeracy (FVR) respondents reported an average job satisfaction of 5.9, and that the high-numeracy respondents reported an average of 5.3.

Table 3 parallels Table 2 but relates to the other column in Clark and Oswald (1996)'s Table 5 — an estimate for satisfaction with pay rather than with the job overall. In this case, increased income is a strong predictor of satisfaction even in naive estimates. On the other hand, higher education again strongly predicts lower satisfaction, after adjusting for household income, in conventional models. This may make sense if the primary effect of education in this context is to set expectations about pay. In any case, for satisfaction with pay, the FVR mixture model corroborates the estimates of the naive ordered logit model.

How can the coefficient estimates remain relatively unchanged in the presence of such a high degree of FVR? While column (5) of Table 3 shows that income and higher education levels predict lower propensity for FVR, and that 31% of respondents used a simplified response scale for answering this question, the net effect of FVR on the estimated coefficients is small. This can be understood by considering the distribution of latent wellbeing values, with reference to the discussion in Section 4.1 and the Remark for Appendix Proposition E.2. For this sample, the number of respondents rounding up from 6 to 7 or from 3 to 4 is balanced by the number rounding down from 2 to 1 or from 5 to

4.²⁸ Appendix Figure F.3 shows the estimated distributions of responses which would have been given in the absence of any FVR (second row), for both job and pay satisfaction. All education levels exhibit broad distributions of latent pay satisfaction, and all carried out some degree of FVR.

6.2. Australia: Powdthavee et al. (2015)

More recently, Powdthavee et al. (2015) have shed some further light on the apparent negative or insignificant returns to education in life satisfaction regressions. They articulate a more considered causal model for the impact of educational attainment on overall life evaluations, taking into account several of the multiple non-monetary channels through which education is expected or known to affect life. In particular, they allow for mediating effects of education through health, marriage, child-rearing, and employment, in addition to income. They conclude that “education is likely to be positively related to overall life satisfaction through many different channels, even when *ceteris paribus*

²⁸ As discussed earlier and as this example shows, there is no simple relationship between the extent of FVR and the size of net biases, due to the possibility of offsetting contributions to bias and the importance of detailed distributional features of the sample. It is also worth noting that the model is capable of accounting for a high fraction of extreme values (1s and 7s, here) as scale boundary effects rather than FVR. Indeed, it is also capable of accounting for a central peak (here, at 4) without appealing to the existence of any FVR. Instead, the model estimate suggests that respondents were simplifying the scale, and that the independently-estimated fractions of respondents who did so were the same (28% and 31%) for the two questions.

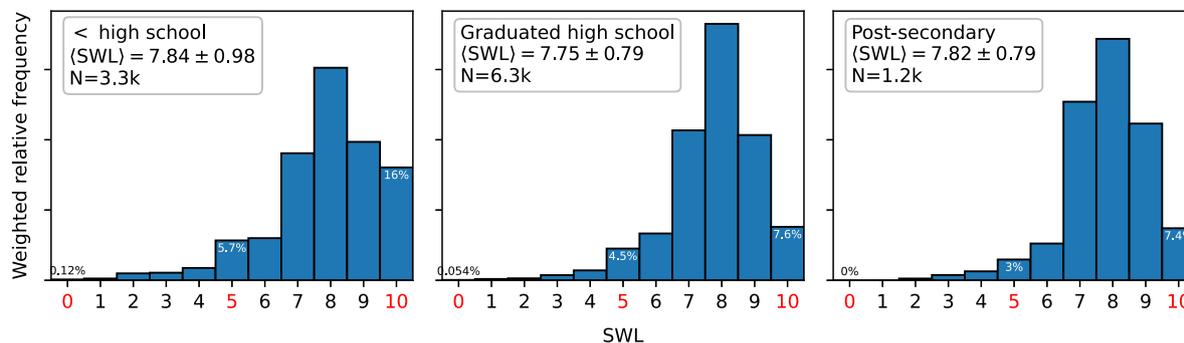


Fig. 5. LS responses in Australia by education level for those over 25 years of age.

Table 4
Estimates of life satisfaction in the 2010 cycle of HILDA.

	Conventional		Mixture model	
	OLS (1)	ologit (2)	FVRI→0 (3)	Mixture (4)
Life satisfaction (β_S)				
log(HH income)	.13 [†] (.019)	.11 [†] (.027)	.10 [†] (.024)	.21 [†] (.026)
educHigh	-.15 [†] (.032)	-.27 [†] (.044)	-.26 [†] (.041)	-.055 (.045)
educCollege	.042 (.047)	.055 (.052)	.054 (.057)	.087 [†] (.059)
constant	6.4 [†] (.21)			
Numeracy (β_N)				
constant				1.47 [†] (.074)
educHigh				.93 [†] (.081)
educCollege				.30 (.15)
FVRI			0 [†] (0)	.11 [†] (.009)
obs.	10744	10744	10744	10744
log likelihood	-19290	-18185	-18211	-18104

education itself has a negative and statistically significant relationship with overall life satisfaction”. Thus, while identifying some positive indirect effects of education, their analysis does not account for the overall negative effect of education on life satisfaction.

Here I do not integrate their panel data mediation pathways into the FVR model, which would go beyond the scope of this paper. Instead, I use one cycle (2010) of the HILDA survey (see Powdthavee et al., 2015) to test the same questions as above: how much of the negative overall association between education and life satisfaction is accounted for by FVR behavior? and how biased is the income coefficient when FVR is ignored?

Fig. 5 shows a familiar pattern in weighted life satisfaction response distributions when separated by education level. Here the focal value enhancements are more subtle, but anomalously high response fractions for 5 and 10 are noticeable at least in the lowest education group, and the proportions of each focal value decrease across education groups.

Table 4 shows the comparison in now-familiar form of the naive estimates of income and education effects on life satisfaction in Australia (columns 1, 2, and 3) with an estimate of the FVR model in column (4). In the FVR-aware model, the coefficient on income approximately doubles, jumping by 4 standard errors. The additional effect of college degree attainment after finishing high school becomes weakly positive, and the effect of high school graduation climbs by 5 standard errors.

6.3. First nations and Métis in Canada

Next I pick on my own prior work by re-examining a paper which reported an anomalously low benefit of income for a sample of Indigenous (First Nations and Métis) peoples in Canada (Barrington-Leigh and Sloman, 2016). In addition to estimating a negative effect of income on life satisfaction, we found an average life satisfaction among Indigenous respondents that was equivalent to that of the general population, despite the stark objective challenges faced by the former groups, including disproportionate levels of discrimination and socioeconomic disadvantage with respect to the rest of the Canadian population. Barrington-Leigh and Sloman (2016) suggested as a possible interpretation that total income is not well measured by the standard income question for this group, but remain “cautious and skeptical” about the data overall.

This case study relates to the importance of being able to use life satisfaction data across diverse cultural and economic circumstances. It also demonstrates the use of the mixture model on a small sample. The data come from two Canadian surveys: the national Equality, Security and Community survey (General ESC, $N = 3725$) and its follow-up small sample of on- (70%) and off-reserve (30%) First Nations and Métis peoples²⁹ in the Canadian Prairies (Aboriginal ESC, $N = 446$). As can be

²⁹ Both of these groups are considered Aboriginal (and, along with the Inuit, Indigenous).

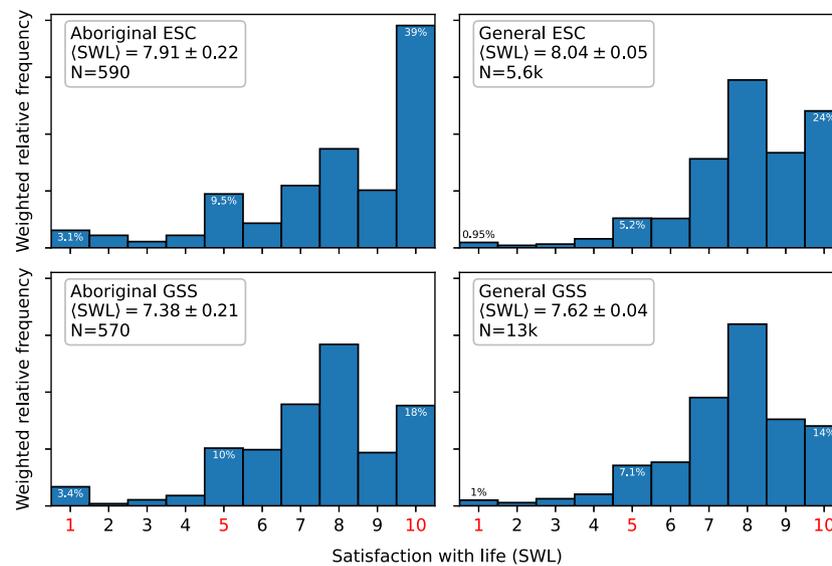


Fig. 6. Life satisfaction of Indigenous Canadians (left panels) and the whole population (right panels). The second row shows similar patterns in the nation-wide General Social Survey. While the Aboriginal ESC is a distinct sample from the General ESC, the panel labeled “Aboriginal GSS” is simply a subset of the full GSS sample.

seen in the first panel of Fig. 6, an enhancement at LS=10 in the Aboriginal ESC sample makes it the modal response value and may go some way to explaining the high mean reported LS. Indeed, this is likely the first report of a LS distribution with such a dominant response at its top value. On the other hand, respondents also gave plenty of 7s, 8s, and 9s, each with higher frequency than LS=5. Below I use the cognitive mixture model to assess how much this distribution might be biased by FVR, and whether the anomalous estimates in Barrington-Leigh and Sloman (2016) are reversed.

The first column of Table 5 shows a conventional ordered logit estimate of 10-point life satisfaction of the Aboriginal sample. For consistency with Barrington-Leigh and Sloman (2016), the education variable is a more continuous variable than in the previous two applications, being measured on ten steps ranging from no primary school to a PhD or professional degree. Once again, and despite a sample size of only 446, a significantly negative coefficient on education shows that, after adjusting for income, those with higher education report lower life satisfaction. In addition, the coefficient on log household income is estimated to be most likely negative, with a 95% confidence interval between -0.40 and $+0.08$.

The second column reports the estimate of a cognitive mixture model. Education strongly predicts numeracy, i.e., use of the full response scale. Most importantly, the education anomaly in the earlier analysis is resolved when FVR is taken into account: the confidently negative education coefficient is replaced by a weakly positive point estimate with a 95% confidence interval between -0.08 and $+0.17$. The weaker anomaly of a negative income coefficient is also partly resolved; in its place is one centered closely on zero with similar precision.

In order to address the surprisingly high average life satisfaction reported by Indigenous respondents, I next use a pooled model to compare groups after controlling for income and education. Pooled estimates of the Canada-wide respondents and the First Nations/Métis sample are shown in Columns (3) and (4) of Table 5. Adjusting for income and education, the Aboriginal respondents report 0.30 higher life satisfaction than non-Aboriginal. Although the explanatory variables here are few and the model is simple, this positive boost is counterintuitive for the reasons described above. However, when FVR is accounted for (Column 4), this situation is reversed, with the Aboriginal respondents reporting a weakly lower life satisfaction than others with similar income and education. In this model, education, income, and Aboriginal status are all allowed to predict FVR behavior. Education and income positively predict lower propensity for FVR behavior, as

expected, while Aboriginal status has the equivalent effect on FVR as a two-point reduction in educational attainment level, for instance from completing a technical or community college certification to completing only high school.

For the pooled sample, the mixture model estimates a 70% higher income coefficient and corrects the strongly negative education effect of the naive model estimate with a weakly positive one.

6.4. Ranking of U.S. states by happiness

The United States is somewhat exceptional in that there are no prominent domestic surveys assessing subjective wellbeing with more than a 4-point response,³⁰ with the exception of the Gallup Daily Poll, which poses the Cantril Ladder question on an 11-point scale.

In this section, I investigate the extent to which a ranking of states by average reported life evaluations is biased by focal-value response behavior. I find that state-level differences in educational attainment relate to state-level differences in FVR. Applying the cognitive mixture model to these data provides a counterfactual “latent” or “corrected” mean life evaluation for each state, allowing for a comparison between a naive ranking and a corrected ranking of states.

Ranking of happiness around the world garners considerable attention, with over one million visits and downloads of the World Happiness Report each year. Below, the USA case demonstrates that a bias in rankings occurs when mean responses are taken at face value.

Fig. 7(a)’s horizontal axis shows the distribution of state mean responses to the Cantril Ladder framing of life evaluation³¹ in the 2019 (final) wave of the Gallup Daily Poll. Counter-intuitively, these means are uncorrelated with the fraction of respondents in each state who provided the answer “10” on the 0–10 scale (vertical axis). Fig. 7(b) gives some suggestion as to why. States with higher high school completion rates have lower tendency to answer “10”. Fig. 7(c) and (d) show an example of how much states can differ in terms of FVR. The weighted response distribution for Mississippi, which has a high incidence of answer “10”, is remarkably different from that of Washington DC,³² with the lowest incidence, even though their mean responses are similar.

³⁰ However, two international datasets, the World Values Survey and the Gallup World Poll, do so on 10 and 11 point scales, respectively.

³¹ See Appendix G.7 for the precise wording of the question.

³² The survey also covered Washington DC. It is aggregated here alongside the states, even though it is entirely urban, unlike any state.

Table 5
Estimates of life satisfaction of Indigenous Canadians.

	Aboriginal sample		Combined sample	
	Conventional ologit (1)	Mixture model Flexible (2)	Conventional ologit (3)	Mixture model Flexible (4)
Life satisfaction (β_S)				
education (10 pt)	-.082 (.041)	.044 (.064)	-.067[†] (.014)	.016 (.019)
log(HH income)	-.16 (.12)	.001 (.15)	.32[†] (.046)	.56[†] (.052)
Aboriginal sample			.30 (.13)	-.11 (.12)
Numeracy (β_N)				
constant		.77* (.43)		2.1[†] (.21)
education (10 pt)		.35[†] (.15)		.35[†] (.056)
log(HH income)				.54[†] (.12)
Aboriginal sample				-.74[†] (.15)
FVRI		.34[†] (.076)		.15[†] (.017)
obs.	446	446	4171	4171
log likelihood	-806	-821	-7409	-7307

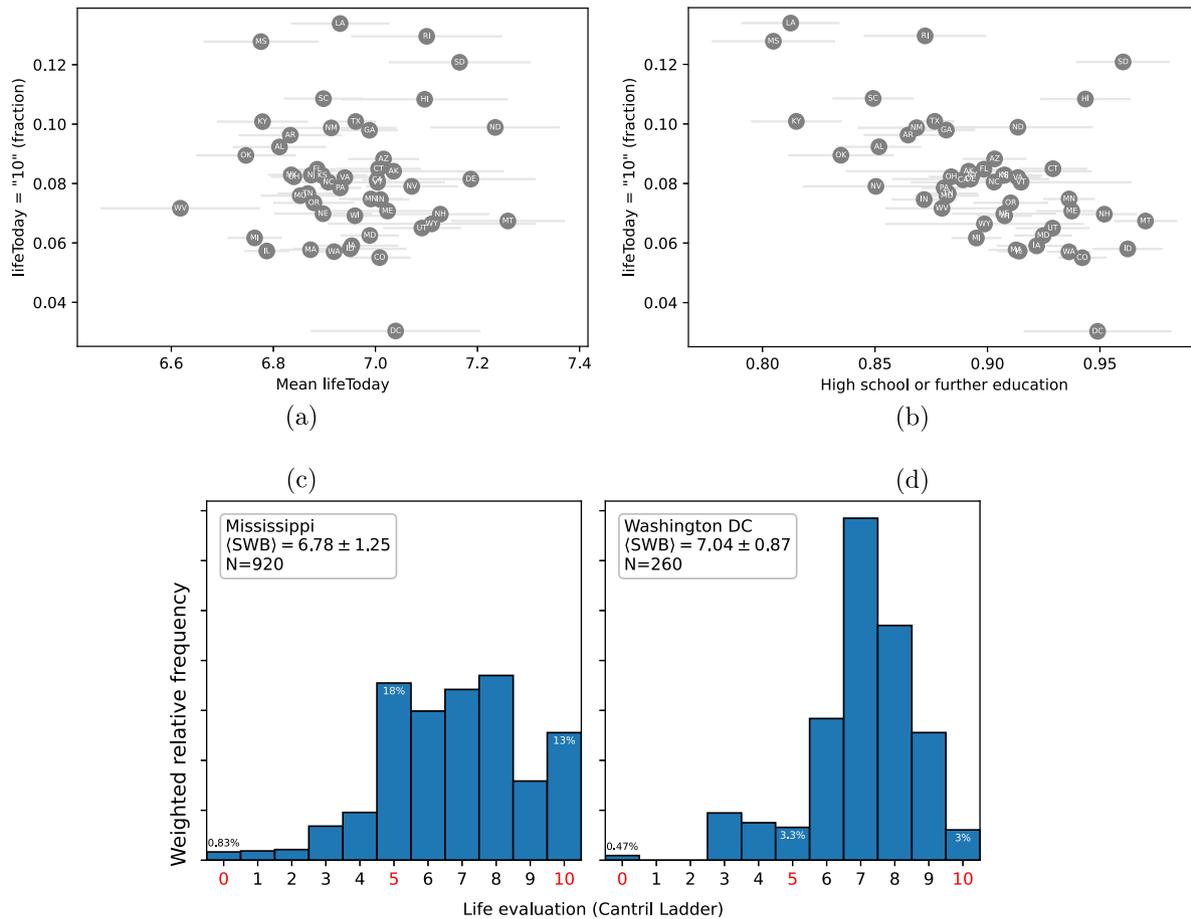


Fig. 7. State-level responses and propensity to answer “10”. Use of focal value “10” is uncorrelated with mean life evaluation (a) but is negatively correlated with state average education (b). Whiskers show standard errors. Panels (c) and (d) show two examples of distributions with and without heavy FVR behavior. See [Appendix Table F.6](#) for further descriptive statistics.

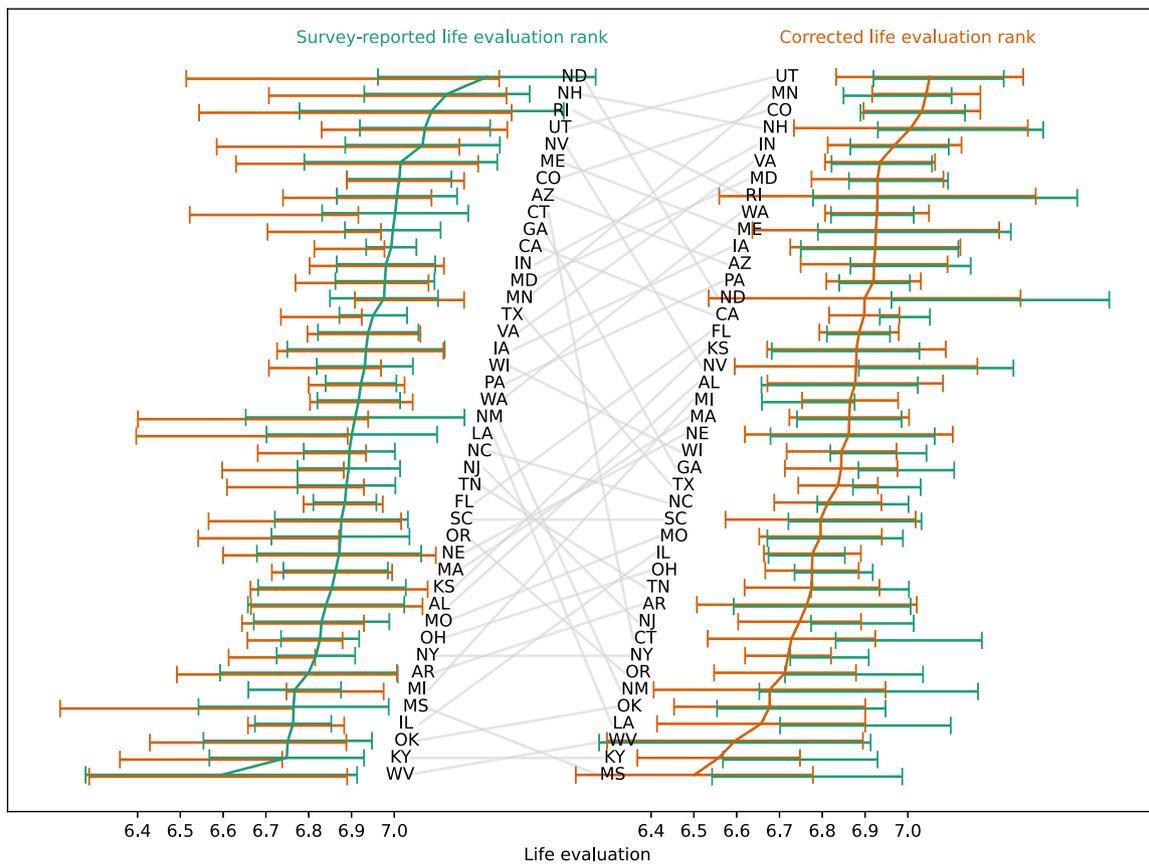


Fig. 9. Observed and corrected U.S. state rankings. Error bars show 95% confidence intervals. States without at least one of each possible response to the SWB question are omitted.

“extreme” values. One natural extension of the model described in this paper is to allow for the inclination to round (FVR) to vary separately for each focal value, effectively creating a mixture of eight “types” in the case of three focal values.

A deeper analysis of panel data will also be important, through an extension to incorporate fixed effects into the model developed in this paper. Preliminary analysis of panel data with a 5-point scale for LS, treating values 1, 3, and 5 as focal values, shows that the probability of LS changing from the middle value is decreasing in education. Traditional 1st-differences approaches for panel fixed effects are invalid because, for instance, the dependence of the 3 → 4 transition is not the mirror of the 4 → 3 transition.

Another extension of the model used in this paper will be to incorporate instrumental variables. Fortunately, this is relatively straightforward in Bayesian estimation frameworks, in which a single-step estimation procedure for instrumental variables is natural, subject to the normal exclusion restrictions (Drèze, 1976; Kleibergen and Zivot, 2003).

As a proof of principle and in light of the descriptive evidence, this paper focuses on the idea of numeracy and on education as a primary predictor of FVR. Understanding the role of secondary influences, such as other demographic variables, fatigue, the cost of time, or motivation with respect to the survey, may help to identify other biases or to design better surveys.

Survey and questionnaire interface design is a further topic of future work. While the present study carries out an *ex post* determination of how respondents have used a subjective numerical scale, it may make sense to give respondents this choice up front. An interactive survey interface could dynamically offer different degrees of precision or resolution in responses, thus accommodating variation in cognitive capacity and other differences in the confidence of respondents’ answers. Open-ended graphical scales may be one means to accomplish

this, but further research into ways to elicit a statement of precision from respondents would be valuable. The potential for creativity and innovation is high, given the increasing availability of technology during an interview.

Depending on one’s perspective, the present findings on response behavior, happiness income coefficients, and mean response biases may be taken as a warning of how difficult it would be to realize the most ambitious implementations of LS as a guide to policy (Frijters et al., 2020; Frijters and Krekel, 2021; Barrington-Leigh and Escande, 2018; Barrington-Leigh, 2016; Happiness Research Institute, 2020; Barrington-Leigh, 2021; UK Treasury, 2021; MacLennan et al., 2021) or, conversely, as another reassuring example of the robustness of LS inference to potential flaws inherent in its cognitive complexity, and possibly even a defense of the rough magnitudes of estimated effects that have become so reproducible in study after study. I take away both of these messages.

Declaration of competing interest

The author declares no conflict of interest.

Data availability

The authors do not have permission to share data.

Acknowledgments

I am grateful for discussion and comments from Fabian Lange, Kevin Lang, Andrew Oswald, Idrissa Ouili, Nadia DeLeon, two helpful referees, co-editor Keith Marzilli Ericson, and audiences in Vancouver, Chicago, Oxford, Cambridge, Montreal, and Winnipeg, among others. This work was supported by Canada’s Social Sciences and Humanities Research Council (SSHRC) grant 435-2016-0531.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jpubeco.2023.105052>.

References

- Alwin, D.F., 1997. Feeling thermometers versus 7-point scales: Which are better? *Sociol. Methods Res.* 25, 318–340.
- Barrington-Leigh, C., 2014. Consumption externalities. In: Michalos, A.C. (Ed.), *Encyclopedia of Quality of Life and Well-Being Research*. Springer, pp. 1248–1252, URL: <http://wellbeing.ihsp.mcgill.ca/?p=pubs#ConsumptionExternalities>.
- Barrington-Leigh, C., 2016. The role of subjective well-being as an organizing concept for community indicators. In: Meg Holden, R.P., Stevens, C. (Eds.), *Community Quality of Life and Wellbeing: Best Cases VII*. Springer, pp. 19–34. <http://dx.doi.org/10.1007/978-3-319-54618-6>, Community Quality-of-Life Indicators.
- Barrington-Leigh, C., 2021. Life satisfaction and sustainability: A policy framework. *SN Soc. Sci.* <http://dx.doi.org/10.1007/s43545-021-00185-8>.
- Barrington-Leigh, C., Escande, A., 2018. Measuring progress and well-being: A comparative review of indicators. *Soc. Indic. Res.* 135, 893–925. <http://dx.doi.org/10.1007/s11205-016-1505-0>.
- Barrington-Leigh, C., Sloman, S., 2016. Life satisfaction among Aboriginals in the Canadian Prairies: Evidence from the Equality, Security and Community survey. *Int. Indig. Policy J.* 7, <http://dx.doi.org/10.18584/iipj.2016.7.2.2>.
- Blanchflower, D.G., Bell, D.N., Montagnoli, A., Moro, M., 2014. The happiness trade-off between unemployment and inflation. *J. Money Credit Bank.* 46, 117–141. <http://dx.doi.org/10.1111/jmcb.12154>.
- Blanchflower, D., Oswald, A., 2004. Well-being over time in Britain and the USA. *J. Public Econ.* 88, 1359–1386. [http://dx.doi.org/10.1016/S0047-2727\(02\)00168-8](http://dx.doi.org/10.1016/S0047-2727(02)00168-8).
- Boes, S., Winkelmann, R., 2006. Ordered response models. *Allgemeines Stat. Archiv.* 1, 167–181. <http://dx.doi.org/10.1007/s10182-006-0228-y>.
- Bradburn, N.M., Sudman, S., Wansink, B., 2004. *Asking Questions: The Definitive Guide to Questionnaire Design — for Market Research, Political Polls, and Social and Health Questionnaires*. John Wiley & Sons.
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: A probabilistic programming language. *J. Stat. Softw.* 76, <http://dx.doi.org/10.18637/jss.v076.i01>.
- Cheung, F., Lucas, R.E., 2014. Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Qual. Life Res.* 23, 2809–2818. <http://dx.doi.org/10.1007/s11136-014-0726-4>.
- Clark, A., Ettilé, F., Postel-Vinay, F., Senik, C., Van der Straeten, K., 2005. Heterogeneity in reported well-being: Evidence from twelve European countries. *Econ. J.* 115, C118–C132. <http://dx.doi.org/10.1111/j.0013-0133.2005.00983.x>.
- Clark, A.E., Flèche, R., Powdthavee, N., Ward, G., 2019. *The Origins of Happiness: The Science of Well-Being over the Life Course*. Princeton University Press.
- Clark, A.E., Frijters, P., Shields, M.A., 2008. Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *J. Econ. Lit.* 46, 95–144. <http://dx.doi.org/10.1257/jel.46.1.95>.
- Clark, A., Oswald, A., 1996. Satisfaction and comparison income. *J. Public Econ.* 61, 359–381. [http://dx.doi.org/10.1016/0047-2727\(95\)01564-7](http://dx.doi.org/10.1016/0047-2727(95)01564-7).
- Conti, G., Pudney, S., 2011. Survey design and the analysis of satisfaction. *Rev. Econ. Stat.* 93, 1087–1093. http://dx.doi.org/10.1162/REST_a_00202.
- Deaton, A., 2008. Income, health and wellbeing around the world: Evidence from the Gallup World Poll. *J. Econ. Perspect.* 22 (53), <http://dx.doi.org/10.1257/jep.22.2.53>.
- Department of Finance, 2021. *Measuring What Matters: Toward a Quality of Life Strategy for Canada*. URL: <https://www.canada.ca/en/department-finance/services/publications/measuring-what-matters-toward-quality-life-strategy-canada.html>.
- Dolan, P., Layard, R., Metcalfe, R., 2011. *Measuring Subjective Well-Being for Public Policy*. Office for National Statistics paper.
- Dolan, P., Peasgood, T., White, M., 2008. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *J. Econ. Psychol.* 29, 94–122. <http://dx.doi.org/10.1016/j.joep.2007.09.001>.
- Drèze, J.H., 1976. Bayesian limited information analysis of the simultaneous equations model. *Econometrica* 44, 1045–1075, URL: <http://www.jstor.org/stable/1911544>.
- Easterlin, R., 1974. Does economic growth improve the human lot? Some empirical evidence. In: David, P., Reder, M. (Eds.), *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz*. Academic Press, pp. 98–125.
- Easterlin, R.A., 1995. Will raising the incomes of all increase the happiness of all? *J. Econ. Behav. Organ.* 27, 35–47.
- Easterlin, R.A., 2013. Happiness, growth, and public policy. *Econ. Inquiry* 51, 1–15. <http://dx.doi.org/10.1111/j.1465-7295.2012.00505.x>.
- Everitt, B.S., 1988. A finite mixture model for the clustering of mixed-mode data. *Stat. Probab. Lett.* 6, 305–309.
- Everitt, B.S., Merette, C., 1990. The clustering of mixed-mode data: A comparison of possible approaches. *J. Appl. Stat.* 17, 283–297.
- Exton, C., Smith, C., Vandendriessche, D., 2015. Comparing Happiness Across the World: Does Culture Matter? OECD Statistics Working Papers, <http://dx.doi.org/10.1787/5jrppz9b2-en>.
- Ferrer-i Carbonell, A., 2005. Income and well-being: An empirical analysis of the comparison income effect. *J. Public Econ.* 89, 997–1019. <http://dx.doi.org/10.1016/j.jpubeco.2004.06.003>.
- Ferrer-i Carbonell, A., Frijters, P., 2004. How important is methodology for the estimates of the determinants of happiness? *Econ. J.* 114, 641–659. <http://dx.doi.org/10.1111/j.1468-0297.2004.00235.x>.
- Frick, J.R., Goebel, J., Schechtman, E., Wagner, G.G., Yitzhaki, S., 2006. Using analysis of Gini (ANOGI) for detecting whether two subsamples represent the same universe: The German Socio-Economic Panel Study (SOEP) experience. *Sociol. Methods Res.* 34, 427–468. <http://dx.doi.org/10.1177/0049124105283109>.
- Frijters, P., Clark, A.E., Krekel, C., Layard, R., 2020. A happy choice: Wellbeing as the goal of government. *Behav. Public Policy* 4, 126–165. <http://dx.doi.org/10.1017/bpp.2019.39>.
- Frijters, P., Krekel, C., 2021. *A Handbook for Wellbeing Policy-Making: History, Theory, Measurement, Implementation, and Examples*. Oxford University Press, <http://dx.doi.org/10.1093/oso/9780192896803.001.0001>.
- Giustinelli, P., Manski, C.F., Molinari, F., 2020. Tail and center rounding of probabilistic expectations in the health and retirement study. *J. Econometrics* <http://dx.doi.org/10.1016/j.jeconom.2020.03.020>.
- Goff, L., Helliwell, J.F., Mayraz, G., 2018. Inequality of subjective well-being as a comprehensive measure of inequality. *Econ. Inq.* 56, 2177–2194. <http://dx.doi.org/10.1111/ecin.12582>.
- Grimes, A., 2021. *Budgeting for wellbeing*. In: *A Modern Guide to Wellbeing Research*. Edward Elgar Publishing, pp. 266–281.
- Hamamura, T., Heine, S.J., Paulhus, D.L., 2008. Cultural differences in response styles: The role of dialectical thinking. *Personal. Individual Differences* 44, 932–942. <http://dx.doi.org/10.1016/j.paid.2007.10.034>.
- Hamilton, K., Helliwell, J.F., Woolcock, M., 2016. Social capital, trust and well-being in the evaluation of wealth. <http://dx.doi.org/10.3386/w22556>.
- Happiness Research Institute, 2020. *Wellbeing Adjusted Life Years: A universal metric to quantify the happiness return on investment*. URL: <https://www.happinessresearchinstitute.com/waly-report>.
- Hasegawa, H., Ueda, K., 2011. Measuring inequality of subjective well-being: A Bayesian approach. *J. Socio-Econ.* 40, 700–708. <http://dx.doi.org/10.1016/j.socce.2011.05.009>.
- Helliwell, J., Barrington-Leigh, C., 2011. How much is social capital worth? In: Jetten, J., Haslam, C., Haslam, S.A. (Eds.), *The Social Cure: Identity, Health, and Well-being*. Taylor and Francis, pp. 55–71, URL: <http://wellbeing.research.mcgill.ca/publications/w16025-and-appendix.pdf>.
- Helliwell, J., Barrington-Leigh, C., Harris, A., Huang, H., 2010. International evidence on the social context of well-being. In: Diener, E., Helliwell, J., Kahneman, D. (Eds.), *International Differences in Well-Being*. Oxford University Press, pp. 213–229.
- Helliwell, J., Putnam, R., 2004. The social context of well-being. *Philos. Trans. R Soc. Lond. B. Biol. Sci.* 359, 1435–1446. <http://dx.doi.org/10.1098/rstb.2004.1522>.
- Helliwell, J., Putnam, R., 2007. Education and social capital. *Eastern Econ. J.* 33, 1–19. <http://dx.doi.org/10.1057/ej.2007.1>.
- Kapteyn, A., van Praag, B., van Herwaarden, F., 1978. Individual welfare functions and social reference spaces. *Econom. Lett.* 1, 173–177.
- Khorramdel, L., von Davier, M., Pokropek, A., 2019. Combining mixture distribution and multidimensional irtree models for the measurement of extreme response styles. *Br. J. Math. Stat. Psychol.* 72, 538–559. <http://dx.doi.org/10.1111/bmsp.12179>.
- Kleibergen, F., Zivot, E., 2003. Bayesian and classical approaches to instrumental variable regression. *J. Econometrics* 114, 29–72. [http://dx.doi.org/10.1016/S0304-4076\(02\)00219-1](http://dx.doi.org/10.1016/S0304-4076(02)00219-1).
- Kroh, M., et al., 2006. *An experimental evaluation of popular well-being measures*.
- Landua, D., 1992. An attempt to classify satisfaction changes: Methodological and content aspects of a longitudinal problem. *Soc. Indic. Res.* 26, 221–241. <http://dx.doi.org/10.1007/BF00286560>.
- Lau, A., Cummins, R., McPherson, W., 2005. An investigation into the cross-cultural equivalence of the personal wellbeing index. *Soc. Indic. Res.* 72, 403–430. <http://dx.doi.org/10.1007/s11205-004-0561-z>.
- Layard, P.R.G., 2011. *Happiness: Lessons from a New Science*. Penguin, UK.
- Levinson, A., 2012. Valuing public goods using happiness data: The case of air quality. *J. Public Econ.* 96, 869–880. <http://dx.doi.org/10.1016/j.jpubeco.2012.06.007>.
- Lewbel, A., 2019. The identification zoo: Meanings of identification in econometrics. *J. Econ. Lit.* 57, 835–903. <http://dx.doi.org/10.1257/jel.20181361>.
- Luttmer, E.F.P., 2005. Neighbors as negatives: Relative earnings and well-being. *Q. J. Econ.* 120, 963–1002. <http://dx.doi.org/10.1093/qje/120.3.963>.
- MacLennan, S., Stead, I., Rowlatt, A., 2021. *Wellbeing Discussion Paper: Monetisation of Life Satisfaction Effect Sizes: A Review of Approaches and Proposed Approach*. Technical Report, UK Government, URL: <https://www.gov.uk/government/publications/green-book-supplementary-guidance-wellbeing>.
- OECD, 2013. *OECD Guidelines on Measuring Subjective Well-Being*. OECD Publishing, <http://dx.doi.org/10.1787/9789264191655-en>.
- Powdthavee, N., 2008. Putting a price tag on friends, relatives, and neighbours: Using surveys of life satisfaction to value social relationships. *J. Soc.-Econ.* 37, 1459–1480. <http://dx.doi.org/10.1016/j.socce.2007.04.004>.

- Powdthavee, N., Lekfuangfu, W.N., Wooden, M., 2015. What's the good of education on our overall quality of life? A simultaneous equation model of education and life satisfaction for australia. *J. Behav. Exp. Econ.* 54, 10–21. <http://dx.doi.org/10.1016/j.socec.2014.11.002>.
- Riddell, A., Hartikainen, A., Carter, M., 2020. PyStan (2.19.1.1). PyPI.
- Saris, W., Van Wijk, T., Scherpenzeel, A., 1998. Validity and reliability of subjective social indicators. *Soc. Indic. Res.* 45, 173–199.
- Senik, C., 2005. Income distribution and well-being: what can we learn from subjective data? *J. Econ. Surv.* 19, 43–63. <http://dx.doi.org/10.1111/j.0950-0804.2005.00238.x>.
- Stan Development Team, 2018. Stan modeling language users guide and reference manual. version 2.18.0. URL: <http://mc-stan.org/>.
- Stevenson, B., Wolfers, J., 2008. Happiness inequality in the United States. URL: <http://www.nber.org/papers/w14220>.
- Stutzer, A., 2004. The role of income aspirations in individual happiness. *J. Econ. Behav. Organ.* 54, 89–109. <http://dx.doi.org/10.1016/j.jebo.2003.04.003>.
- Stutzer, A., Frey, B.S., 2006. Does marriage make people happy, or do happy people get married? *J. Socio-Econ.* 35, 326–347. <http://dx.doi.org/10.1016/j.socec.2005.11.043>, the Socio-Economics of Happiness.
- Uebersax, J.S., 1999. Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Appl. Psychol. Meas.* 23, 283–297. <http://dx.doi.org/10.1177/01466219922031400>.
- UK Treasury, 2021. Wellbeing Guidance for Appraisal: Supplementary Green Book Guidance. Technical Report, UK Government, URL: <https://www.gov.uk/government/publications/green-book-supplementary-guidance-wellbeing>.
- Van Praag, B., Kapteyn, A., 1973. Further evidence on the individual welfare function of income: An empirical investigation in The Netherlands. *Eur. Econ. Rev.* 4, 33–62.
- Weng, L.J., 2004. Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educ. Psychol. Measur.* 64, 956–972. <http://dx.doi.org/10.1177/0013164404268674>.
- Wu, M., 2005. The role of plausible values in large-scale surveys. *Stud. Educ. Eval.* 31, 114–128. <http://dx.doi.org/10.1016/j.stueduc.2005.05.005>, measurement, Evaluation, and Statistical Analysis.