Embedding as a pitfall for survey-based welfare indicators: Evidence from an experiment

Freie Universität Berlin

1. Introduction

Figure 1: OECD Better Life Index weighting tool
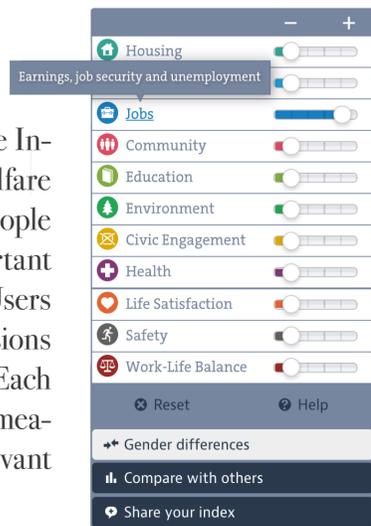
The OECD Better Life Index is an interactive welfare measure allowing people to decide what is important for their wellbeing. Users weight eleven dimensions on a scale from 0 to 5. Each dimension embeds measurable indicators relevant for quality of life.

2. The embedding effect

The valuation of an item might differ depending on whether it is valued on its own or as part of a more inclusive item - the so-called **embedding effect** (Kahnemann and Knetsch, 1992).

This motivated us to investigate the validity of the OECD BLI. In the case of the OECD BLI, users assign weights to over-arching well-being dimensions instead of their underlying indicators. Thus, the valuation of a dimension might differ when allowing for an individual rating of the underlying indicators.

3. The experiment

We replicated the OECD's weighting tool ('R-BLI') in an experimental website, inviting participants to rate the different dimensions of quality of life.

THE EXPERIMENT
- R-BLI website was accessible from: 18/01/16 – 12/02/16, using a ticket n° (ID)
- Distribution of 2,730 flyers : display of web-address and ID n°
- Distribution in undergraduate lectures across German universities: Rostock, Berlin (TU, FU), Magdeburg, Göttingen, Bochum, Wuppertal, Dresden, Frankfurt
- Response rate of 19.7% (number of observations: 538)

Participants were assigned to five different versions of the R-BLI weighting tool that vary the *Jobs* dimension.

The Jobs dimension embeds four indicators:
- job security
- personal earnings
- long term unemployment rate
- employment rate

Figure 2: Treatment and control versions of the weighting tool

control group 1	treatment group 1	treatment group 2	treatment group 3
Exact replication of OECD Better Life Index	We split up jobs into two dimensions: job quality (JQ) reflecting the indicators personal earnings / job security and labour market (LM) reflecting unemployment.	We withdrew the indicators reflecting Labour market, leaving the indicators for job quality only.	We withdrew the indicators reflecting job quality, leaving the indicators for labour market only.

Adding up the relative weights *(RW)* the treated assign to the two new dimensions should not differ from the controls' relative weighting of the whole Jobs dimension.

Finding RW_{C1} *(Jobs)* $< RW_{T1}$ *(JQ)* $+ RW_{T1}$ *(LM)* instead would indicate regular embedding.

The stronger notion of perfect embedding implies that the maximum of the relative weights T1 participants assign to *JQ* or to *LM* exactly equals the control groups' relative weight of *Jobs*.

4. Results

The average relative weights of the eleven dimensions of control group C1 are highly correlated with those of Germany-based users (Pearson's r = 0.94, Spearman's ρ = 0.92) and those of all worldwide users (r = 0.95, ρ = 0.94).

Thus, we believe that using a replication of the BLI weighting tool did not influence the rating process in a way different from the original.

Figure 3: Relative dimension weights across different user groups
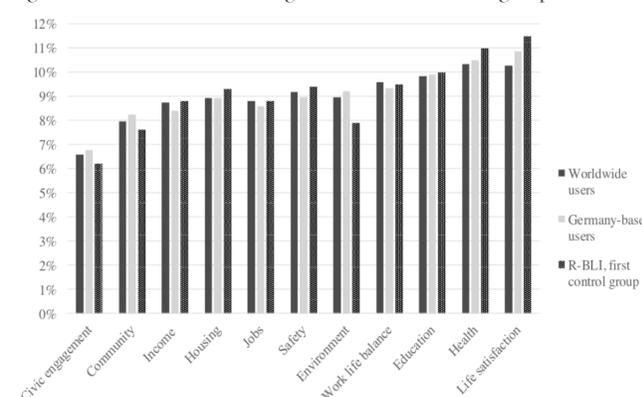
Note: The bars display average relative weights of dimensions according to the ratings of all worldwide users (left bars, 133,920 observations) and all Germany-based users (central bars, 11,897 observations) of the original BLI. These reference data were downloaded from http://www.oecdbetterlifeindex.org/bli/ on 16 January 2018. Right bars display the relative weights based on the ratings of the first control group (105 observations).

We compared the mean relative weights assigned to the Jobs dimension and its treatment 1 replacements Labour Market and Job Quality.

We always use the parametric t-test and the non-parametric Mann-Whitney-U test.

As the following figure illustrates, the treated assign a substantially and significantly higher relative weight to Labour Market and Job Quality in sum (0.141) than the controls attribute to Jobs (0.088). This implies strong embedding effects.

Withdrawing indicators (treatments 2 and 3) does not affect the users valuation of jobs at all.

Figure 4: Dimension ratings across different user groups
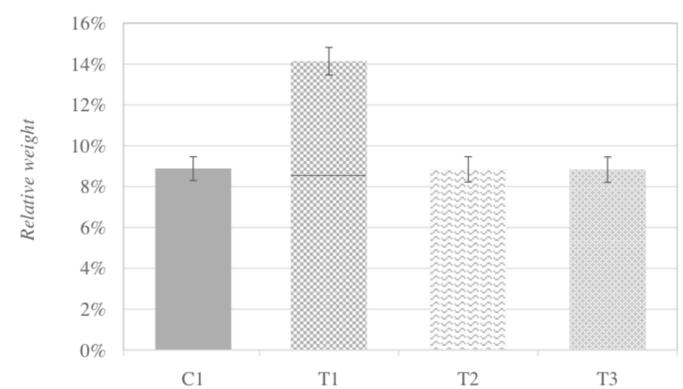
Note: For C1 as well as T2 and T3 the bars display average relative weights of the Jobs dimension. The T1 bar illustrates the sum of average relative weights of the dimensions Job Quality (below the horizontal line) and Labour Market (above the horizontal line). Whiskers denote 95% confidence intervals.

To consider demographic and further information in the identification of embedding effects, we conduct OLS regression analyses, controlling for gender, age, the size of hometown and people's major. Further, we control for variables related to the weighting process, namely previous knowledge of the BLI, time spent weighting and having accessed extra information. The regression analysis confirms the results of the mean analysis.

We moreover conduct subgroup analyses. It turns out that even people who spent a relatively long time weighting, or accessed the extra information, show embedding effects that are of similar size compared to the whole sample across all of the treatments.

5. Conclusion

Our experiment shows that embedding the indicators in dimensions decisively affects people's weightings, that is we observe strong embedding effects. The Better Life Index hence fails to measure citizens' true preferences. Embedding thus opens a gateway to manipulating survey-based welfare measures.

References

Kahneman, Daniel and Jack L. Knetsch (1992a): "Valuing Public-Goods - the Purchase of Moral Satisfaction", Journal of Environmental Economics and Management 22(1), pp. 57–70.

Clemens Hetschko, Freie Universität Berlin, IAB, CESifo
Louisa von Reumont, Freie Universität Berlin
Ronnie Schöb, Freie Universität Berlin, CESifo, ifo Dresden

Please send your comments to Clemens.Hetschko@fu-berlin.de