Comments on: **Adjusting for Scale-Use Heterogeneity in Self-Reported Well-Being** (+ some wider thoughts)

Caspar Kaiser

Warwick Business School

March 16, 2026

What the Paper Does

- Proposes a new framework for adjusting for **scale-use heterogeneity** in self-reported wellbeing data.
- Core ideas: (i) ask respondents *calibration questions* about which perceptions are assumed to be shared. Observed response differences then identify differences in scale use. (ii) Identify both shifts and stretches in scale use.
- Four assumptions:
  1. **Affine transformation:** Scale-use of person A is a positive affine transformation of person B's.
  2. **Common perception:** Calibration questions perceived identically across respondents.
  3. **Response consistency:** Translation functions are the same across questions.
  4. **Independent errors.**

Main Empirical Results

- Correcting for scale-use differences has only modest effects on regression coefficients (Table 6). If this holds broadly, it is good news for the field.
- Correcting for scale-use increases the extent to which subjective reports predict objectively observable quantities (height, weight, pollution).
- After correcting, mean wellbeing is higher and variance is smaller. *What does that really mean?*
- Open questions:
  - Do these results generalise cross-culturally? (MTurk US sample.)
  - Do they hold in traditional surveys (SOEP, LISS, Gallup)?

Practical Implications and Next Steps

- **Sensitivity analysis 1:** How sensitive are estimates (e.g. Table 6) to varying the included set of calibration questions? If results are stable, this suggests the common perception assumption is approximately met.
- **Sensitivity analysis 2:** What happens in other data? What happens in in LMICs? What happens in response to *interventions*?
- **Panel data:** If the "shifter" varies over time and with covariates, individual fixed effects do not solve the problem. $\rightarrow$ Extension to FE models would be nice.
- **Accessibility:** Stata/R/Julia/Python packages would be nice!
- **Relationship to HOPIT:** The two approaches are not nested and make substantively different assumptions.

Other Problems for Wellbeing Measurement

Scale-use heterogeneity (what Benjamin et al. address) is one of several concerns. At least three others:

1. **Linearity:** Even if all respondents use the scale the same way, we don't know if the mapping from underlying states to responses is linear. Non-linear transformations can especially affect relative magnitudes (Bond & Lang 2019; Kaiser & Lepinteur 2025).
2. **Neutrality:** For many applications (e.g. comparing life-saving vs. life-improving interventions), we need to know the *neutral point*.
3. **Correct concept:** Does the argument to the reporting function, u_{it}, equal (flow) utility u_{it}^*?