

## Social Science Measurement in Diverse Societies

Scoping Paper prepared by Jan R. Böhnke, University of York

Prepared for the call:

"Big Data Ideas – Political Perspectives and Radical Statistics" (Bibby & Evans, RadStat)

The movement of evidence-based policy has seen an increased use of data produced by questionnaire assessments. Since the first wave of the "Programme for International Student Assessment" (PISA, 2000) or the "Trends in International Mathematics and Science Study (TIMSS, 1995) this has been especially noteworthy for the realm of educational testing (e.g., PIRLS, 2001; PIAAC, 2011). The debate over the differential efficacy and effectiveness of schools, curricula, educational systems etc ensues in the open and unstopped. But in the UK especially sectors dealing with health and well-being have been quick in the uptake, connecting important assessments and evaluations to questionnaire data. The NHS for example uses a number of questionnaires, patient reported outcome measures (PROMs) to "[...]assess the quality of care delivered to NHS patients from the patient perspective. Currently covering four clinical procedures, PROMs calculate the health gains after surgical treatment using pre- and post-operative surveys." (<http://www.england.nhs.uk/statistics/statistical-work-areas/proms/>) The areas this is applied to are hip replacements, knee replacements, groin hernia and varicose veins. Services of the "Improving Access to Psychological Therapies" initiative (IAPT, <http://www.iapt.nhs.uk/>) use questionnaires to screen patients for eligibility and to monitor patients' progress in brief low-intensity treatments. Services in either setting can be monitored and compared on national level with regards to their relative performance rates.

Interesting as they are, these topics would only provide an overall background for the actual example that this chapter would deal with. An application of questionnaire assessments with far wider social consequences is the UK Government's agenda on wellbeing. It rests heavily on the assessment of the subjective states of respondents in nation-wide surveys as an indicator of overall success of implemented policies ("Measuring National Well-being", ONS). The current preferred methodology for this are still collections of items, that either individually or as a collection (often dubbed "scale") are assumed to measure a construct such as "well-being".

The assessment of well-being in the UK general population is not necessarily a Big Data issue. Collections of items are clearly circumscribed numbers of variables and even if we collected data on each individual living human being in the UK, by standards of some disciplines that would hardly constitute a "large" data set. But it is an example of the wider ranging debate on how policy could and should be evaluated.

As the reader of this is probably aware, there has been wide-spread hesitation to use questionnaire based evaluations outside the domain of cognitive assessment to base hard decisions on. Many of the reservations regarding such assessments are still correct, others are either the result of a lack of motivation to apply high quality (and admittedly expensive) research methodology or they mirror a completely out-dated perception of how questionnaires are built and validated.

This chapter would use the following assumptions, work down the following questions and touch as debates:

- The reliable and valid assessment of a subjective variable like well-being with a collection of questions is possible. The statistical methods and theoretical frameworks for this have been developed in the 1990ies and refined in the past 10 years.
- Current research and government policy is at odds with this.
- The main problem of the applied set of methods deals with the validity of the assessment. This aspect is strongly connected to the notion of inequality: Research has shown and is showing that the interpretation of different items used in these surveys is different across ethnicities and regions (i.e., different types of "well-being" are assessed and comparisons on the raw scores are not informative about true differences between regions or ethnicities) or even across time (e.g., the assessment of well-being in the Health Survey for England between 2010 and 2013 differs in the use of the items, e.g., trend cannot be interpreted directly).
- Consequences for the debate about well-being would have to be explored – this is what I am currently working on so I cannot really make a clear argument, what this would be. On a technical level this would be a call for sustained and topical research investment into such areas if this was identified as a governmental priority. Practically, this would highlight that the differences between regions and the interpretation of trends over time is much more difficult than currently presented, maybe even close to useless (enjoy a look at this: <http://www.ons.gov.uk/ons/interactive/uk-experimental-subjective-well-being---dvc34/index.html>)
- One aspect of this exploration would surely be whether the power differential between the agenda setters and the survey respondents (most of whom are not really close to the government or the ONS) might lead to obvious shifts in results (probably in favour of the government) – although this would be exciting, I sincerely doubt it: I think nobody has enough knowledge on that side of the fence to do this.
- Is the assessment of well-being as an indicator relevant for policy? A question I am not sure I would want to deal with in such a rather technical chapter, but definitely something I would be happy to explore with an interested co-author.