

Can we trust survey data?

Provocative findings from survey data are a daily matter; we see them almost every day in the media, be it in the newspapers, in magazines, in the radio or in television. Most interest is paid to country-wide election studies; we desire to know who is leading in the campaigns and which candidates will likely become a member in the parliament. Similar interest arises world-wide every three years when the OECD publishes the first findings of the PISA.

When reporting the conclusions from important cross-national surveys that are used every year for hundreds of scientific articles in the leading journals, we trust that the data are conducted at very high standards of social science research. This includes that the respondents are selected at random, that they give accurate responses, that in telephone or face-to-face interviews the interviewers conduct their interviews according to the prescribed protocol, and that the employees of the research institutes perform their tasks as stipulated in the contract with their client. In the literature there is a large discussion on how to draw the sample, how to contact the respondents, how to formulate the questions, and many other topics to increase the quality of data. Despite the best efforts of the research community, it is well-known that respondents often provide sub-optimal answers as a way to simplify their task and minimize their efforts to complete the interview. Yet very seldom has the research community focused on the case where respondents give answers that are essentially independent of the content of the questions. Behavior such as this may result in respondents giving the same response to all questions of an item battery, for example, strongly agree. Further, there is very little investigation on fabricated interviews by the interviewers and almost no discussion on data that is fabricated by employees from the research institutes, for example, by copy-and-paste to attain the contracted number of interviews.

We emphasize the term “task simplification” for describing all forms of strong deviations from the expected research norms: respondents giving convenient answers to a set of questions, (almost) independent of the content of the question, interviewers deviating from the prescribed interview protocol, which includes faking, or partially faking, their interviews, , and employees of research institutes fabricating interviews via copy-and-paste, which results in identical responses to large segments of the data.

Common to all three sources of task simplification (respondent, interviewer, and staff of research institutes) is that they minimize the time and effort necessary to complete interviews, or to realize the required sample size and to increase the response rate. Of course, all three sources of task shortcuts also reduce the quality of the data. Our paper examined the cross-national quality of the reports from principals of schools participating in the 2009 PISA.

Since the PISA school-level data are obtained via self-administrated questionnaires, only two sources of simplification are possible: satisficing respondents and dishonest employees in research institutes. We introduced and employed several statistical procedures which help identify the prevalence of likely data fabrication. To illustrate our findings, we show the frequencies of identical

responses to a battery of school climate items in two countries: the USA and Slovenia:

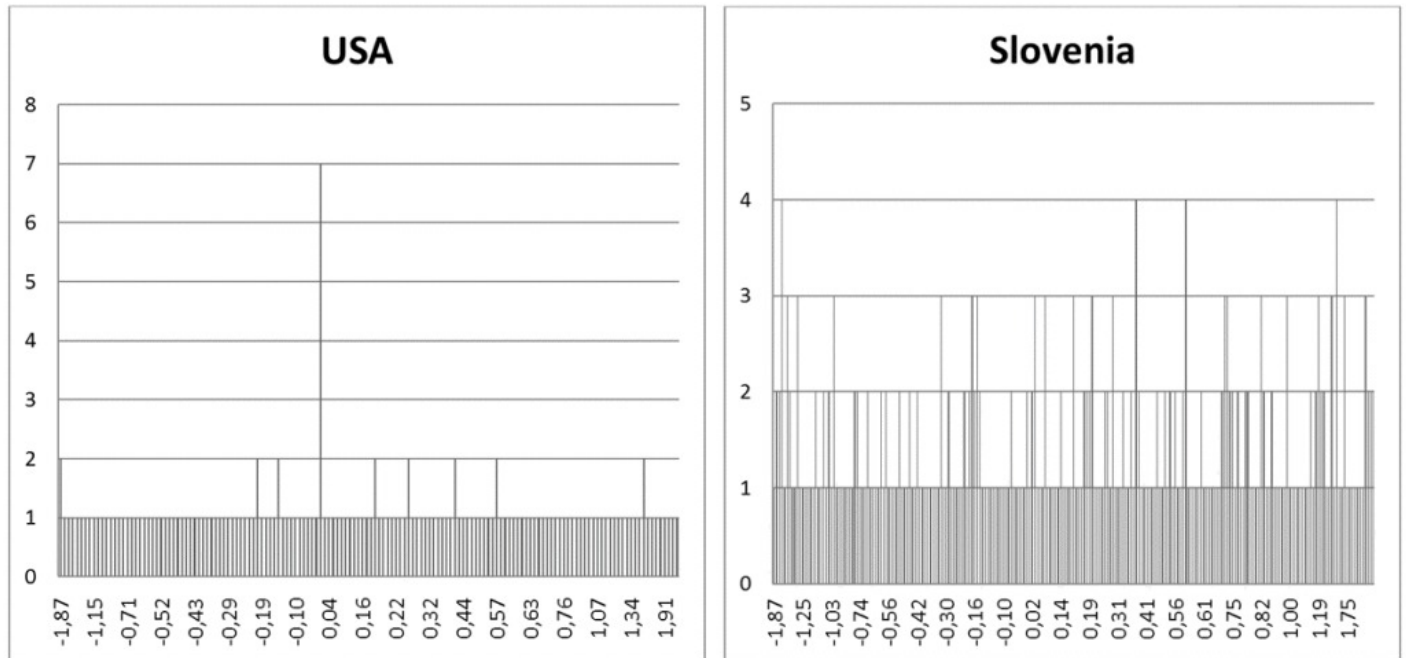


Fig. 1. Barcharts of school climate items, factor scores, USA and Slovenia

Clearly, in the USA almost all response patterns are unique, with only five instances of duplicated patterns (and one instance of seven identical patterns). In contrast, in Slovenia we find many instances of duplicate patterns, a sizable number of triplicates, and four different instances where the same pattern occurs four times. This suggests the prevalence of fraudulent research behavior.

Publication

[Should we trust survey data? Assessing response simplification and data fabrication.](#)

Blasius J, Thiessen V
Soc Sci Res. 2015 Jul