Social Science Research 52 (2015) 479–493

Contents lists available at ScienceDirect

Social Science Research

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

Should we trust survey data? Assessing response simplification and data fabrication

Jörg Blasius [a,*], Victor Thiessen [b,1]

[a] University of Bonn, Institute of Political Science and Sociology, 53113 Bonn, Lennéstr. 27, Germany
[b] Dept. of Sociology and Social Anthropology, Dalhousie University, Halifax, NS B3H 1P9, Canada

ARTICLE INFO

Article history:
Received 5 February 2014
Revised 21 February 2015
Accepted 19 March 2015
Available online 30 March 2015

Keywords:
Data quality
Data fabrication
Undifferentiated responses
Identical response pattern

ABSTRACT

While many factors, such as unit- and item nonresponse, threaten data quality, we focus on data contamination that arises primarily from task simplification processes. We argue that such processes can occur at two levels. First, respondents themselves may engage in various response strategies that minimize their time and effort in completing the survey. Second, interviewers and other employees of the research institute might take various shortcuts to reduce their time and/or to fulfill the requirements of their contracts; in the simplest form this can be done via copy-and-paste procedures.

This paper examines the cross-national quality of the reports from principals of schools participating in the 2009 PISA. We introduce two measures of data quality to document that extreme response simplification characterizes the behavior of substantial numbers of school principals in numerous countries. Additionally, we discovered strong evidence of data fabrication in several countries.

1. Introduction

It is a truism that survey data are vulnerable to numerous sources of error, some of which can be minimized through stringent quality controls by survey research institutes. Perhaps for this reason, researchers increasingly rely on well-known and reputable data sets such as the International Social Survey, the European Social Survey, some National Election Studies, and well-known panel data such as the PSID. Another important data set that generates much attention in the news media and in politics is PISA, the Program for International Student Assessment. The PISA surveys are known to be executed by well-qualified research groups, with stringent technical quality control mechanisms implemented by the Organization for Economic Cooperation and Development (OECD). Part of the data consists of school-level information provided by the principals of the participating schools. In addition to the strict quality controls and the high educational levels of the respondents (principals), one might expect this data to be of exceptionally good quality. On the other hand, the very fact of the importance of the PISA surveys may lead to dynamics that undermine their quality. The purpose of this paper is to examine this issue using the 2009 school-level data, which are freely accessible (http://pisa2009.acer.edu.au/downloads.php).

Our investigation focuses on two components of data quality. First, we search for signs of respondent task simplification behaviors. Here we identify statistically improbable response patterns on sets of items such as the nature of the school climate. Second, we look for evidence of duplicated data that might have been fabricated simply through copy-and-paste

* Corresponding author. Fax: +49 (0) 228 738430.
E-mail addresses: jblasius@uni-bonn.de (J. Blasius), victor.thiessen@dal.ca (V. Thiessen).
[1] Fax: +1 902 494 2897.

http://dx.doi.org/10.1016/j.ssresearch.2015.03.006
0049-089X/

procedures. For these two tasks, we compare each case with all others *within* each country and *between* all countries. To accomplish this efficiently, we apply principal component analysis (PCA) and multiple correspondence analysis (MCA) to obtain factors scores for each respondent. We employ these techniques solely for the purpose of examining the frequency distributions of factor scores to identify multiple occurrences of identical response patterns; i.e., unlike the traditional use of these techniques, we are not interested in possible substantive interpretations.

2. The quality of survey data

Most previous research on poor data quality due to measurement error focused on respondents. Two main approaches were developed in this regard: response styles (Baumgartner and Steenkamp, 2001) and "satisficing" strategies (Krosnick and Alwin, 1987; Krosnick, 1991, 1999). These are briefly summarized next, followed by a review of the more scant literature on the role of interviewers and research institutes.

Oskamp (1977: 37) defined response styles as "systematic ways of answering which are not directly related to the question content". A large literature has identified various response tendencies such as acquiescence, selecting the extreme options, mid-point responding, limited response differentiation, and random or arbitrary responding (Aichholzer, 2013; Baumgartner and Steenkamp, 2001; Van Rosmalen et al., 2010; Watkins and Cheung, 1995). Controlling for response styles is particularly important for cross-cultural studies, since response styles have been found to vary by both race and culture (Bachman and O'Malley, 1984; Dayton et al., 2006; Hamamura et al., 2008; Hui and Triandis, 1989). Including a latent style factor increased the number of countries with factorial invariance (Cambré et al., 2002; Diamantopoulos et al., 2006; Khorramdel and von Davier, 2014) and produced fewer counter-intuitive findings (Elliott et al., 2009; Weech-Maldonado et al., 2008).

Krosnick and others popularized the term "satisficing" to provide a theoretical rationale to explain various response tendencies (cf. Krosnick, 1991, 1999; Krosnick and Alwin, 1987). Krosnick's application is based on Tourangeau's (Tourangeau and Rasinski, 1988; Tourangeau et al., 2000) four-step cognitive process model for producing high-quality information: the respondent must (1) understand the question, (2) retrieve the relevant information, (3) synthesize the retrieved information into a summary judgment, and (4) choose a response option that most closely corresponds with the summary judgment. Satisficing theory emphasizes that it is the limited cognitive ability of some respondents that induces them to engage in satisficing behavior. The empirical evidence is consistent with this emphasis, since educational attainment and cognitive skills are consistent predictors of various indicators of satisficing (Converse, 1976; Krosnick, 1991; Marsh, 1986, 1996; Meisenberg and Williams, 2008; Thiessen and Blasius, 2008; Wilkinson, 1970).

Krosnick et al. (1996) distinguished between weak and strong satisficing. Weak satisficing entails executing all four stages described by Tourangeau, but being less than thorough in doing so. In contrast, strong satisficing omits the retrieval and judgment steps altogether, with respondents simply selecting what they consider to be reasonable responses. An extreme form of "strong satisficing" might consist of selecting only a single response option (such as "strongly agree") to all items in a large set of items. A related form is to consistently choose the most positive answer, especially when it is the socially most desirable one, and hence easily defended (Krosnick et al., 1996: 32). Such behavior results in undifferentiated response patterns (URPs), also known as straight-lining. Respondents producing any kind of URP simultaneously simplify their task and save time. Applying Tourangeau's model, there is minimal need to understand the question or to retrieve relevant information and as a consequence, little information that requires a synthesis into a summary judgment and without a summary judgment, there would be no response option that would most closely correspond to it.

We emphasize task simplification rather than satisficing responses for several reasons. While we agree that cognitive limitation can spawn satisficing behaviors, our concept of task simplification is more general. It permits us to understand satisficing dynamics that occur not only because of high cognitive demands, but also because of lack of interest or knowledge of the topics being investigated (Blasius and Thiessen, 2001). Further, it is also more parsimonious in that it recognizes that all response styles also simplify the task for the respondent. Rather than having to weigh numerous response options, only subsets of them need to be considered. It is arguably for this reason that response styles are, just like manifestations of satisficing, consistently negatively associated with education (Bachman and O'Malley, 1984; Elliott et al., 2009; Greenleaf, 1992; Watson, 1992). Perhaps most importantly, it alerts the researcher that it is not only respondents who might wish to minimize their time and energy commitment: interviewers and employees in research institutes might have the same motivation. For interviewers this can take the form of deviating from the prescribed interview protocol, and includes faking, or partially faking, their interviews (Winker et al., 2013). For example, they can skip time-consuming questions such as long item batteries with plausible fabricated responses filled in later (Blasius and Thiessen, 2013). Under the header of interviewer motivation, cheating was discussed as early as the mid-forties, (Crespi, 1945, 1946; Durant, 1946). Crespi (1945: 431) argued that cheating "lies as much in the structure of the ballots and the conditions of administration as in the personal integrity of the interviewer". He identified questionnaire design features that might demoralize interviewers such as unreasonable length, apparent repetition of questions, as well as complex questions. Another reason for task-simplifying rule-breaking could be "pressure from a supervisor or even higher up" (Nelson and Kiecker, 1996: 1110).

Likewise, staff of research institutes can reduce their effort by fabricating interviews through copy-and-paste procedures. As Nelson and Kiecker (1996) stated for the interviewers, higher-level employees might pressure their subordinates to be more efficient; that is, to process more cases in less time while simultaneously maintaining a high response rate. Greater