Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ

XiaoChi Zhang a,*, Lars Kuchinke b, Marcella L. Woud a, Julia Velten a, Jürgen Margraf a

a Mental Health Research & Treatment Center of Ruhr-Universität Bochum, Germany
b Experimental Psychology, Ruhr-Universität Bochum, Germany

Abstract

Self-report inventories enable efficient assessment of mental attributes in large representative surveys. However, an inventory can be administered in several ways whose equivalence is largely untested. In the present study, we administered thirteen psychological questionnaires assessing positive and negative aspects of mental health. The questionnaires were administered by four different data collection methods: face-to-face interview, telephone interview, online questionnaire, and offline questionnaire. We found that twelve of the questionnaires differed in survey methods. Although, some studies showed that social desirability tends to be highest for telephone survey and lowest for web survey. Furthermore, the effects of social desirability should be the same for the online and offline samples. However, there were no statistically significant differences between the face-to-face and telephone samples for the anxious scale, the stress scale, and the tradition scale. We also found that for eight scales, the online sample was statistically different from the offline sample in the respondent answers. Moreover, the survey method effects were only moderated by age. Finally, measurement invariance across the four survey methods was tested for each self-report measure. There was full strong measurement invariance established for nine of thirteen scales and partial strong measurement invariance for the remaining four scales across the four survey methods. These findings indicated that measurement invariance was affected by different survey methods.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Self-report measures are widely used to study and assess personality characteristics and various aspects of health and behavior. More recently, however, traditional paper pencil surveys have been challenged by computer supported surveys. Since the rapid expanding of the internet, online surveys became more and more popular (Griffiths, Lewis, Ortiz de Gortari, & Kuss, 2014). There are a number of advantages for this approach: simplified work for the interviewers, fast data processing, and low costs (Beebe, Mika, Harrison, Anderson, & Fulkerson, 1997; Rosenfeld, Booth-Kewley, & Edwards, 1993). Not surprisingly, however research found that different survey methods can lead to different responses although the same questions were asked (Kiesler & Sproull, 1986). This is called “mode effect”, and a number of such effects have been identified. Social desirability is one of the most studied mode effects. The results of these studies, however, have been inconsistent. To illustrate, many studies examined data quality and the effects of social desirability when using different survey methods. In some studies, computer surveys yielded similar results as paper and pencil surveys, e.g., on attitude questionnaires (Booth-Kewley, Edwards, & Rosenfeld, 1992) or for personally sensitive questions (Knapp & Kirk, 2003). In other studies, however, different results were found when using different survey methods, e.g., on satisfaction-dissatisfaction questions (Dillman et al., 2008) or on questions about consumption frequency and preferences related to wine (Szolnoki & Hoffmann, 2013). Furthermore, response biases for telephone interviews and internet questionnaires caused by social desirability have been reported (Chang & Krosnick, 2009). Here, more social desirability was manifested for telephone compared to Internet surveys, respectively. Some studies also showed that biases related to social desirability tended to be highest for telephone surveys and lowest for web surveys (Holbrook, Green, & Krosnick, 2003; Kreuter, Presser, &...
More recently, however, a meta-analysis concluded that social desirability was the same in offline, online and paper surveys (Dodou & de Winter, 2014). Hence, this shows that the scientific state concerning the effects of social desirability is still inconsistent, and more research is needed to advance our understanding of its effects and underlying mechanisms.

A possible explanation of these inconsistencies could be the lack of large representative population samples with sufficient power to detect relevant effects. Moreover, in-depth investigations of measurement invariance across different assessment modes are sparse. Some studies examined the measurement invariance when using web surveys compared to paper and pencil methods (Davidov & Depner, 2011; Fang, Wen, & Prybutok, 2014). Human value scales were found scalar invariant between online and paper-pencil surveys in Davidov and Depner’s study. But, in Fang’s study, paper-pencil survey was found nonequivalent to social media surveys on personal and global innovativeness scales. To the best of our knowledge, there is no research yet examining the measurement invariance for psychological questionnaires across common survey methods within representative samples. When comparing groups, it is assumed that the used measures target the same construct in all groups. If this assumption does not hold, however, the comparisons across the groups can neither be evaluated meaningfully nor interpreted adequately. Therefore, the establishment of measurement invariance is a prerequisite when applying self-report measures (Millof & Fischer, 2010). Hence, its investigation is an important target when using self-report measures.

Within this context there is another issue to consider. That is, it may make a difference whether the self-report scales target more or less general, innocuous personality characteristics or more sensitive constructs such as positive or negative aspects of mental health. The latter concepts are often related to issues that many people consider socially sensitive, e.g., social support, represented by the number of friends one has, or personal (un-)happiness (Fydrich, Sommer, Tydecks, & Brähler, 2009; Kessler et al., 2015; Maercker et al., 2015). Following this, our study addressed these particular domains.

The present study had two main foci, namely examining the role of social desirability for and the existence of measurement invariance in various data collection methods assessing positive and negative aspects of mental health. Therefore, four survey methods in four German representative samples were applied: face-to-face interviewing, online questionnaires, offline questionnaires, and telephone interviewing. All four survey methods included thirteen different measures assessing positive and negative mental health. In order to ensure sufficient statistical power and generalizability of the results, we studied large representative population samples (N > 2000 for each sample). There were three research aims. The first is related to the role of social desirability. Social desirability was operationalized as the difference in responses for different kinds of self-report measures for all four survey methods. There were two research questions: Will the largest difference in responses for the different kind of measures occur between online and telephone samples (see Holbrook et al., 2003), or between offline and telephone samples (see Dodou & de Winter 2014)? Will the online sample deliver the same responses for different kind of self-report measures as the offline sample? This would be in line with results of the meta-analysis by Dodou and de Winter (2014). The second aim involved an exploratory question and concerned the moderating role of age, gender, and education level for the observed effect of social desirability. The third aim concerned the measurement invariance. Here, we tested the configural invariance, weak invariance, and strong invariance across the four survey methods.

### 2. Methods

Participants were recruited within the Bochum Optimism and Mental Health Studies (BOOM) program, which aimed to identify protective factors related to positive mental health in different countries. Four representative German samples were tested in 2012, each one using a different data collection method: face-to-face interview, online questionnaire, telephone interview, or offline-panel (Forsa.Omninet). Each sampling had its own procedure:

The face-to-face sample (N = 1870) and the online sample (N = 2039) were both conducted via the market research company GfK, and included the same weighting factors, i.e., age, gender, state, city size, size of household and occupation of head of household. The face-to-face sample used the Computer Assisted Multimedia Questioning (CAM) method and the online sample used the Computer Assisted Web Interviewing (CAWI) method.

The Offline sample (Forsa.Omninet) (N = 2021) was collected by a German market research company named Forsa Ltd. The respondents answered the questions on their home PC or on their TV screen, which are linked to Forsa’s own proprietary environment using a device called “set-top-box”, implying that the internet was not needed for this data collection method. The Forsa.Omninet sample currently consists of 10,000 representatively selected households in Germany. The data was weighted by age, gender, federal state, and education.

The telephone sample (N = 2007) was conducted by another German market research company called USUMA. The sampling frame, which is called “ADM-Telefonstichproben-System”, is based on the amount of available telephone numbers in Germany as updated by the government agency in charge of the German telephone network. It covers all possible telephone numbers in Germany, independent of whether they are used or not. The data was weighted by age, gender, and household size.

All these specification of weighting factors are based on the most recent data provided by the federal statistical office in Germany.

#### 2.1. Positive mental health scales

##### 2.1.1. Sense of coherence

This scale is a shortened form (Schumacher, Gunzelmann, & Brähler, 2000) of the 29-item-version from Antonovsky (Antonovsky, 1987) and consists of 9 items assessing comprehensibility, manageability, meaningfulness. Each item (e.g. ‘Do you have the feeling that you are in an unfamiliar situation and don’t know what to do?’) has a 7-point Likert scale. This short version was validated by Schumacher in a representative German sample. Cronbach’s α in our four samples varied from 0.78 to 0.89.

##### 2.1.2. Resilience

This scale is a shortened form (Schumacher, Leppert, & Gunzelmann, 2004) of the 25-item-version from Wagnild and Young (Wagnild & Young, 1993). It consists of 11 items assessing positive resilient personality characteristics on a 7-point Likert scale from 1 (‘I disagree’) to 7 (‘I agree’). The German version has been validated by Schumacher et al. Cronbach’s α in our four samples varied from 0.88 to 0.93.

##### 2.1.3. Satisfaction with life

This scale (Diener, Emmons, Larsen, & Griffin, 1985) consists of 5 items focusing on global life satisfaction. A 7-point Likert scale from 1 (‘strongly disagree’) to 7 (‘strongly agree’) indicates the agreement with each item. Cronbach’s α in our four samples varied from 0.84 to 0.92.
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات