Personality and Individual Differences 111 (2017) 281–290

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Reliability and completion speed in online questionnaires under consideration of personality

Christopher Harms [a,*], Lina Jackel [a], Christian Montag [b,c]

[a] Department of Psychology, University Bonn, Bonn, Germany
[b] Institute for Psychology and Education, Ulm University, Ulm, Germany
[c] Key Laboratory for NeuroInformation, Center for Information in Medicine, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China

ARTICLE INFO

Article history:
Received 17 November 2016
Received in revised form 3 February 2017
Accepted 5 February 2017
Available online xxxx

Keywords:
Personality
Impulsiveness
Reliability
Online questionnaire
Completion time

ABSTRACT

While self-administered web questionnaires are increasingly used in psychological and market research, the quality of collected data has to be continuously inspected. Two key questions were investigated in this paper: Does the overall completion speed in an online self-report questionnaire influence the reliability of personality scales? We assessed the Five-Factor Model of Personality (using NEO-FFI) and Impulsiveness (using BIS-11). Second, we asked if Impulsiveness predicts the overall completion speed.

In total, 532 participants (436 females, 96 males; mean age = 25.57 years) engaged in an online study to answer these questions. Replicating previous findings, no difference in the reliabilities was found for fast or slow respondents. While underlining the effect of Age on the completion speed, our data indicated evidence against our hypothesis of an influence of Impulsiveness on completion time using a Bayesian approach. Similar results could be observed using classical inferential methods. Of note, no effect could be observed for the Five-Factor Model of Personality and completion time either. Therefore, personality traits are not associated with individual differences in completion time in our investigated sample. We discuss our findings in a broader context of survey research and give a perspective for future research opportunities.

1. Introduction

Online questionnaires are becoming a *de facto* standard in psychological or sociological research, market research and many other scientific and non-scientific domains. This standard offers several advantages in data collection and analysis such as being able to avoid missing data or errors in transferring information from a paper-pencil document to an electronic data file. Also, larger and more representative samples can be obtained through sampling online sources and panels. Online participant pools such as Amazon Mechanical Turk (MTurk) have helped to collect samples for several studies and some authors have argued that those samples are more representative when compared to traditional samples in psychological research (Buhrmester, Kwang, & Gosling, 2011; Rand, 2012; Rouse, 2015). Online survey platforms also allow designing questionnaires with more complex logic (e.g. filters to direct participants to different questions based on their previous responses) or other adaptive strategies for testing.

While exploiting the benefits, less consideration goes into the question of data quality in practice. Past research has shown the equivalence of pen-and-paper and online application of personality questionnaire extensively in many different areas (Campos, Zucoloto, Bonafé, Jordani, & Maroco, 2011; Carlbring et al., 2007; Chuah, Drasgow, & Roberts, 2006). In contrast to traditional data collection, additional data on the participant and its response behavior is available. So-called "meta-data" on the user (e.g. used web-browser, operating system) and "para-data" (data on the process of filling in the questionnaire, e.g. reaction time, scrolling distance on the page, typing speed, etc.) can be very easily and objectively collected (Stieger & Reips, 2010) without interfering with the actual questionnaire. This allows researchers to include further behavioral and technical variables in their analysis, such as precisely measured response latencies (which would belong to the category of "para-data"). These meta- and para-data can be used as a first indicator of survey data quality (Furnham, Hyde, & Trickey, 2013; Gummer & Rossmann, 2015; Heerwegh, 2003). In particular, when using any para-data to filter respondents in a sample, it is crucial for survey operators or researchers to understand the implications on both the quality of their data and on the validity of conclusions. Filtering respondents might otherwise lead to a bias in the sample and can reduce generalizability of the results. Thus, it seems a reasonable question to ask if para-data used to filter respondents such as response latencies are

* Corresponding author at: Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Psychologie, Kaiser-Karl-Ring 9, 53111 Bonn, Germany.
E-mail address: christopher.harms@uni-bonn.de (C. Harms).

http://dx.doi.org/10.1016/j.paid.2017.02.015
0191-8869/

correlated with any of the constructs, which again are connected to the theory investigated. If, for example, response times in an online marketing survey were negatively correlated with household income, filtering respondents based on their individual response time and removing very fast participants would result in a sample with under-representation of households with low income. If household income was related to the research question in any direct or indirect way, conclusions based on the filtered data have to account for this bias. If not household income but personality dimensions were correlated with response time, the effect on the research question might be less direct but still relevant.

From a psychometric point of view, the question of data quality is often investigated in terms of reliability. For scores in psychometric questionnaires, internal consistency is commonly used as a measure for reliability through indexes such as Cronbach's α (Cronbach, 1951) or McDonald's ω (McDonald, 1999). Montag and Reuter (2008) have shown that reliability (in terms of internal consistency) is not affected by the time participants needed to complete a questionnaire. Their question is a different perspective of the questions raised above: Participants in online questionnaires might just “click through” a questionnaire without taking much effort to answer the questions or they might interrupt the questionnaire and do something unrelated before returning to the questionnaire. Researchers might, thus, be inclined to remove participants from their sample as they expect low data-quality from those participants. Prior results, however, show that a psychometric perspective on data quality seems not to be affected by the respondents' completion speed. As the design by Montag and Reuter (2008) required the researchers to assess completion time manually based on emails sent by the server, the measurement might have been imprecise. The present study uses automatically generated time-stamps stored on the server to measure overall completion time of the questionnaire. This reduces possibilities for inaccuracies and biases in the analysis. Moreover, as the previous study investigated possible effects of completion time on internal consistencies of the Affective Neuroscience Personality Scales (Davis, Panksepp, & Normansell, 2003), the present study aims at the extension of findings to other prominent self-report questionnaires often used in personality psychology – namely the NEO-FFI and the Barratt Impulsiveness Scale (BIS-11).

Regarding the use of para-data in an online study, it is noteworthy that no consensus has been reached on the question how long it typically takes to complete a certain number of items and what factors play an important role to predict completion time. An answer to these questions is of tremendous importance, because it would help to understand if any bias is introduced through using a cut-off time to remove participants or, on the other hand, if such a cut-off time can reasonably be used to discard data as invalid. While the socio-demographic variables Age and Educational level have been previously shown as influences on overall completion time (Yan & Tourangeau, 2008), other influences might also be reasonable to assume: As reading and comprehension of questions and answers is required, cognitive ability is likely to affect completion time in any questionnaire (Maschke, 1989; Voas, 1957). In a different study, when asked about their attitudes, the stability of these attitudes were related to the participants' response time (Bassili, 1993; Bassili & Fletcher, 1991; Heerwegh, 2003). Additional cognitive processes that are required to answer questions might also increase the time needed to complete a questionnaire. For instance, faking (i.e. giving answers to represent a certain profile that does not match the true attitude of the respondent) has been shown to affect the time participants need to complete questionnaires (Holden, 1995, 1998; Holden & Hibbs, 1995; Komar, Komar, Robie, & Taggar, 2010). In general, different cognitive processes take place when someone is answering a self-report questionnaire. Thus, it is reasonable to expect a number of different predictors for completion time. Besides demographics, ability or faking, personality might also play a role to understand how participants answer a questionnaire and thus how long they need.

Beyond the replication of the earlier results by Montag and Reuter (2008), a further research question in the present study covers this potential influence of personality on the overall completion time of self-report questionnaires. In particular, we are interested how the self-chosen time rhythm in filling in questionnaires is related to personality: In our research scenario, participants are not hurried or pressured to fill in the inventories in a given time window. Since the completion of a questionnaire requires reading and understanding instructions and items, one might expect cognitive ability to influence completion time. Educational level is, therefore, included in our study as a rough proxy for cognitive ability as in earlier (and somewhat similar) research on Impulsiveness and completion time (Gummer & Rossmann, 2015; Malloy-Diniz, Fuentes, Borges Leite, Correa, & Bechara, 2007; Reeve, 2007; Yan & Tourangeau, 2008).

For personality, the NEO-FFI is one of the most commonly used questionnaires when research focuses on the Five-Factor Model of Personality. It consists of 60 self-report items, each related to one of the five factors, namely Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness (Costa & McCrae, 1992), that have been described and investigated extensively in personality research (Borkenau & Ostendorf, 1993; Costa & McCrae, 1992; Egan, Deary, & Austin, 2000; Körner et al., 2008; Whiteside & Lynam, 2001). While the NEO-FFI focuses on higher order personality dimensions, the BIS-11 (Patton, Stanford, & Barratt, 1995) assesses a more specific part of human behavior, namely Impulsiveness, which is assessed using 30 items.

Impulsiveness is a lower-order personality dimension and has different conceptualizations and relationships to the Five-Factor Model (for a thorough overview see e.g. Whiteside & Lynam, 2001). Miller, Joseph, and Tudway (2004) have reviewed different theoretical constructs and operationalizations and identified three components of Impulsiveness which they labeled (1) “non-planning and dysfunctional impulsive” behavior, (2) “functional venturesomeness” and (3) “reward responsiveness and drive”. Dickman (1990) and Reeve (2007) have highlighted the influence of functional Impulsiveness in tests of cognitive processes and mental ability. Relating to the first of these components, Patton et al. (1995) have constructed Impulsiveness as orthogonally to anxiety and revised the Barratt Impulsiveness Scale. Given the many different measures for different concepts of Impulsivity, we selected the Barratt Impulsiveness Scale in its current version (BIS-11; Patton et al., 1995) for the present study as one of the most commonly used measures in the field. It proposes three subtraits of Impulsiveness, namely Attentional Impulsiveness, Motor Impulsiveness and Non-Planning Impulsiveness. Motor Impulsiveness assesses the tendency for “acting without thinking” (Stanford et al., 2009, p. 386). Non-Planning Impulsiveness focuses on the “lack of […] forethought” in decision making (Stanford et al., 2009, p. 386). Finally, the third subfacet, Attentional Impulsiveness, covers both attentional and cognitive instability, that is the “inability to focus attention or concentrate” (Stanford et al., 2009, p. 386). While the subfacets cover different aspects of the dysfunctional Impulsiveness construct, they are intercorrelated with correlations between 0.39 and 0.50 (Stanford et al., 2009, p. 388, Table 3). The global score, thus, represents an indicator for a global, underlying tendency towards “non-planning and dysfunctional impulsiveness”.

In general, it seems reasonable to assume an effect of Impulsiveness on the response behavior in questionnaires following the presented rationale: dysfunctional Impulsiveness “as a predisposition toward rapid, unplanned reactions to […] external stimuli without regard to the negative consequences of these reactions […]” (Moeller et al., 2001, as cited in Stanford et al., 2009, p. 385) should also play a role in self-report situations where questions are presented as external stimuli. As prior research has shown, impulsive subjects are, for example, faster in reaction-time experiments (Edman, Schalling, & Levander, 1983) and slower in reactions to a Stroop paradigm (Enticott, Ogloff, & Bradshaw, 2006). For speeded tests of cognitive ability Reeve (2007) highlighted the importance of functional Impulsiveness. With respect to response times for self-report questionnaires, in which participants do not have