The Effect of Face-to-Face Interviewing on Personality Measurement

Luisa Hilgert, Martin Kroh, David Richter

DIW SOEP

The Effect of Face-to-Face Interviewing on Personality Measurement

Luisa Hilgert, Martin Kroh, David Richter

DIW Berlin

Abstract

In recent years, an increasing number of nationally representative surveys in the social sciences and economics have implemented the Big Five model of personality. While many personality inventories were originally developed in the context of self-administered questionnaires, they are often used by large surveys in face-to-face interview settings instead. Drawing on an experimental research design, we studied the effect of this switch in the method of data collection on measurement invariance as well as measurement error and interviewer effects in the Innovation Sample of the German Socio-Economic Panel (SOEP-IS). Although in some cases we found slightly stronger associations between interviewer and respondent personality in face-to-face settings, the results generally suggested strict measurement invariance–and therefore full comparability–across methods of data collection.

Keywords: Personality, Big Five, ESEM, Face-to-Face, Interviewer Bias

The Effect of Face-to-Face Interviewing on Personality Measurement

Introduction

In recent years, personality measurements have enjoyed increasing popularity outside of psychological research, particularly in the social sciences and economics (e.g., Heineck, 2011; Uysal and Pohlmeier, 2011). Studies pioneering the use of personality measures include the Socio-Economic Panel Study (SOEP), the Household, Income and Labour Dynamics in Australia (HILDA) Survey, and the British Household Panel Survey (BHPS). While psychological research traditionally uses self-administered questionnaires to measure personality, many surveys in the social sciences and economics use face-to-face interviews instead. However, little is known about the effect of this mode-switch on the measurement of personality.

A large body of survey research documents that the method of data collection can indeed have an impact on the quality of survey responses (for a meta-analysis, see Leeuw, 1992). Whereas face-to-face interviews enable the use of complex instruments and help to prevent misunderstandings, self-administered interviews often ensure greater privacy and fewer reactivity problems such as socially desirable responding (i.e., the tendency of respondents to give positive self-descriptions and to seek approval from the interviewer in their survey responses). Previous investigations of method effects in personality measures either were restricted to sub-populations (Sawhney & Cigularov, 2014) or used non-experimental data (Lang, John, Lüdtke, Schupp, & Wagner, 2011) to compare these two methods. In the present experimental study based on a nationally representative sample of the German population, personality measures were collected through the use of (1) face-to-face interviews and (2) self-administered questionnaires (with an interviewer present in the household). In the basic setup, respondents were randomly assigned to one of the two groups, which differed only in the methods used for the collection of data on personality. First, we investigated (strict) measurement equivalence, which is required to compare personality scores and reliability in the two modes. Specifically, we compared measurement and structural invariance across both interviewer-administered and self-administered interviews. Second, we studied interviewer bias as previous research has shown that the presence of interviewers and specific interviewer characteristics, such as gender, race, and political attitudes, can influence response behavior (for an overview, see Schaeffer, Dykema, and Maynard, 2010). We expect the same to hold for interviewer personality: Since respondents tend to seek approval from interviewers as a form of impression management, they may adjust their responses in personality tests to fit their perceptions of the interviewer's personality. We considered both the variance in responses attributable to interviewers as well as the extent to which interviewer personality correlates with respondents' answers on the personality questionnaire.

Materials and Methods

Sample. The SOEP is a large, random household sample that has been conducting annual interviews since 1984. The SOEP currently has around 28,000 adult respondents across Germany. The SOEP Innovation Sample, SOEP-IS (Richter & Schupp, 2012), provides an additional longitudinal sample for particularly innovative research projects. Both SOEP and SOEP-IS implement the Big Five constructs using a short version known as the BFI-S (Big Five Inventory-SOEP, Gerlitz and Schupp, 2005). In the 2013 version of the SOEP-IS study, half of a total of 5,101 participants answered the BFI-S in the standard computer-based face-to-face method (**C**omputer **A**ssisted **P**ersonal **I**nterviewing, CAPI), while the other half answered the questions through a self-administered questionnaire during the personal interview (**C**omputer **A**ssisted **S**elf **I**nterviewing, CASI). In the CAPI group, the interviewer read the questions and answer options given on the computer screen aloud to the respondents. The respondents then stated their answers and the interviewer recorded the responses on the computer. In the CASI group, interviewers were asked to interrupt the standard face-to-face interviewing at a given point during the interview for the section on personality measurement. Interviewers then turned the screen toward the respondents, who were asked to read the corresponding instructions and complete their answers on their own. In the CASI, the interviewers did not leave the households. However, they were asked not to intervene in the self-administered parts of the interview. Due to the longitudinal character of SOEP-IS, it is important to note that 1,606 of the participants in 2013 had answered the BFI-questionnaire in previous years, typically in the face-to-face mode.

Participants who did not answer all of the BFI-S questions in 2013 (1.4%) were excluded from the analyses. In the face-to-face group, 2,490 respondents provided complete data for the BFI-S, while in the self-administered condition complete data was available for 2,511 participants. There were no differences with respect to nonresponse (χ^2-test: p=0.723) between the two study groups. There were no statistically significant differences in terms of the gender composition (χ^2-test: p=0.295) or family status distribution (χ^2-test: p=0.396). Mean age, however, was slightly higher in the face-to-face (M=52.0) than in the self-administered (M=50.6) group (t-test: p<.01).

The mode of data collection was randomized at the household level, so that the same mode of data collection was applied to all household members. The interviews were conducted by a total of 324 interviewers. The median number of interviews carried by interviewers was 13. Almost all interviewers conducted both face-to-face interviews as well as self-administered interviews. Additionally, the information on the personality of 217 interviewers was available from interviewer questionnaires that were conducted in 2006 and 2012.

Material: The BFI-S. Personality was measured with a 16-item German short version of the Big Five Inventory (BFI-S: Gerlitz and Schupp, 2005; validity: Hahn, Gottschling, and Spinath, 2012). The five-factor model of personality identifies the traits of openness (O), conscientiousness (C),

extraversion (E), agreeableness (A), and neuroticism (N) (McCrae & Costa, 1997). All personality factors, with the exception of openness (for which a four-item scale was used) were measured with three items on a 7-point Likert scale that ranged from '1' (does not apply at all) to '7' (applies completely).

As for the consistency of the respective items, McDonalds ω (McDonald, 1999) was 0.65 for openness, 0.59 for conscientiousness, 0.67 for extraversion, 0.45 for agreeableness, and 0.63 for neuroticism. The interpretation of the ω values is the same as of the commonly used Cronbachs α.[1] The moderate consistency values reflect the few items per dimension. As to the stability of BFI mean scores, we drew on the longitudinal nature of personality measurement in the SOEP-IS. More specifically, we inspected test-retest correlations of the factor mean scores across four years. For the participants that had answered the BFI-S questionnaire before, the retest correlation was $r = 0.57$ for openness, $r = 0.46$ for conscientiousness, $r = 0.58$ for extraversion, $r = 0.41$ for agreeableness, and $r = 0.57$ for neuroticism.[2] The moderate stability found for these values can be attributed to the long test-retest time interval of four years or more.

Procedures for inspecting measurement invariance.   Since the assumption of the absence of cross-loadings in classical confirmatory factor analysis was too restrictive for our data, we instead applied Exploratory Structural Equation Modeling (ESEM; Asparouhov & Muthén, 2009) using quartimin rotation. Exploratory structural equation modeling is an integration of exploratory and confirmatory factor analysis. It provides a tool for multigroup comparisons of means and covariances of the latent factors (see below: Cross-Group Structural Equivalence) as well as for multigroup comparisons of factor loadings, intercepts and residual variances (see below: Cross-Group Measurement Equivalence). As Horn's Parallel Analysis (Horn, 1965) indicated a five-factor structure for both methods of data collection, we extracted five factors in our ESEM-Analysis.

Procedures for inspecting interviewer bias.   We employed hierarchical linear modeling of the factor scores in order to decompose variance between interviewers and respondents nested within interviewers. Moreover, we used hierarchical modeling to study the effects of interviewers' personalities on respondents' personality measures.

---

[1] RMSEAs and model comparisons based on the Akaike information criterion (AIC) and likelihood ratio tests for goodness of fit suggested the use of congeneric models, which imply that all items meant to belong to a common factor measure the same phenomenon. Cronbachs α would have relied on essential tau-equivalence.

[2] For openness, the test-retest correlation was significantly lower in the face-to-face ($r = 0.578$) than in the self-administered group ($r = 0.595$). This was surprising insofar as the stability was slightly lower when the method of data collection did not change. A tentative interpretation may be that there was a higher reliability of measurements obtained through the self-administered personality measures.

Results and Discussion

Invariance Across Modes of Assessment

Cross-Group Measurement Equivalence. To assess whether personality traits were measured equally in both groups, we tested for measurement equivalence. We followed a stepwise procedure, adding restrictions and comparing models[3]. The model with strict invariance (equal factor loadings as well as intercepts and variances of residuals in both groups) was not rejected on a 1% α-level (χ^2 diff. test against strong invariance: p=0.034). Hence, we can safely assume strict measurement equivalence across both groups, i.e., scores calculated from the items belonging to the same factor are comparable across groups.

Cross-Group Structural Equivalence. To assess if the latent factor means and covariances between factors were the same across groups, we compared a model that set these restrictions to a model that did not. The model with strict invariance plus equal factor means and equal factor covariances was not rejected (χ^2 diff. test against strict invariance: p=.336). Hence, we can also assume structural equivalence across both groups, i.e., reliabilities were the same in the two methods of data collection.

Big Five Factor Structure

At this point we knew that we had a five-factor structure in our data and that this structure was invariant in both self-administered and interviewer-administered questionnaires. To assess if the items related to the five-factor structure in the anticipated way, as stated by the BFI-S, we inspected the item loadings of the strong measurement equivalence model. Only one out of 16 convergent loadings was lower than 0.62. Thirteen convergent loadings were greater than 0.7. All but two divergent loadings were of very small magnitude ($< .30$). In sum, we concluded that the anticipated Big Five structure is present in our data, irrespective of mode of data collection.[4]

Interviewer bias

To assess potential interviewer bias, in a first step the variance of the factor scores was decomposed in hierarchical linear models nesting respondents within households and within interviewers. Intraclass correlation coefficients (ICC) for interviewers explained relatively little variance (O: 5.5%, C: 6.4%, E: 1.9%, A: 4.9%, N: 1.4%)[5] These results were very similar to previous studies, showing that there is a higher susceptibility of the constructs of conscientiousness, agreeableness and

---

[3]Details in the online appendix.

[4]All loadings are documented in the online appendix.

[5]Due to the regional allocation of interviewers, the ICC estimates were presumably upward-biased. The lack of variation between interviewer and region precludes the estimation of cross-classified models disentangling regional and interviewer ICC.

openness to interviewer bias (Paulhus, 2002). More importantly, we found no significant differences in ICCs across self-administered and interviewer-administered modes of data collection. The lack of differences in interviewer effects between modes could be due to the interviewers being present in the households in both modes.

Table 1

*The Association of Interviewer Personality and Respondent Personality*

Dependent var.	Model I Respondent Openness	Model II Respondent Conscientiousness	Model III Respondent Extraversion	Model IV Respondent Agreeableness	Model V Respondent Neuroticism
Constant	4.93 **	5.84 **	4.90 **	5.44 **	3.80 **
CASI (ref. CAPI)	-0.02	-0.01	0.03	0.01	-0.01
Interviewer O	0.10 **				
CASI × interv. O	-0.07 *				
Interviewer C		0.08 *			
CASI × interv. C		0.00			
Interviewer E			0.02		
CASI × interv. E			-0.05		
Interviewer A	0.04	0.14 **	0.06	0.11 **	0.00
CASI × interv. A	-0.02	-0.10 *	0.02	0.02	-0.01
Interviewer N					0.01
CASI × interv. N					0.01
σ^2_{Int}	0.05	0.05	0.03	0.04	0.01
σ^2_{HH}	0.20	0.09	0.05	0.06	0.03
σ^2_{Res}	0.92	0.73	1.27	0.81	1.62

Note: Interviewer bias is inspected using ML hierarchical regression. Measures are based on a pooled dataset including both modes. Scores of the participants are regressed on a constant, the mode, the centered interviewer's score on the respective personality dimension, the centered interviewer's score on agreeableness, and interaction terms of the mode and the interviewer's characteristics. Interviewer characteristics are centered. O: openness, C: conscientiousness, E: extraversion, A: agreeableness, N: neuroticism.

* $p < .05$; ** $p < .01$

Figure 1. Association between interviewers' and respondents' personalities is moderated by
mode of assessment.
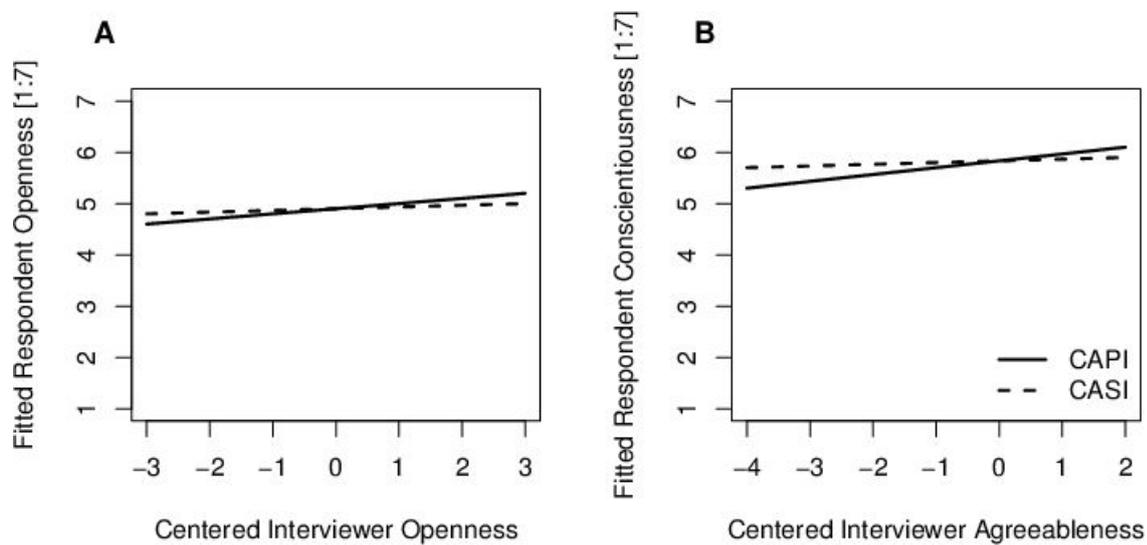
In a second step, respondents' mean personality scores were regressed on the personality scores
of their respective interviewer. In fact, the interviewers' scores on openness ($p<.01$), conscientiousness
($p<.05$), and agreeableness ($p<.01$) turned out to be statistically significant predictors of the respective
interviewees' scores (see Table 1). In addition, interviewers' agreeableness significantly predicted
interviewees' conscientiousness ($p<.001$). These results suggest that there was a certain degree of social
pressure to conform due to the mere presence of the interviewer. However, the effects were of very
small magnitude, the largest effect being a one-point increase in interviewer agreeableness associated
with a 0.14-point increase in respondent conscientiousness in the face-to-face group.

In two cases, we found slightly stronger interviewer effects in the interviewer-administered
face-to-face method: Figure 1 plots the positive associations between interviewers' and participants'
openness (1A) as well as interviewers' agreeableness and participants' conscientiousness (1B).
Compared to the face-to-face group, the correlations were significantly smaller in the self-administered
group.

Conclusions

In applied research, the mode of collecting data on personality traits differs widely across
studies. While many of these differences may be considered innocuous, such as the differences between
paper-based and computer-based self-reports, interviewees may react strongly to a face-to-face
interview setting. To test mode effects, we randomly allocated respondents to self-administered versus
interviewer-administered measurements of Big Five personality items in the context of the SOEP
Innovation Sample (SOEP-IS). Although we found that two correlations between interviewers' and
respondents' personality traits were slightly lower in the self-administered than in the face-to-face

interview, interviewer effects were generally low and comparable across modes. Moreover, exploratory structural equation modeling analysis showed that strict measurement invariance holds across the two groups. This means that scores calculated from items belonging to the same dimension are reasonably comparable across modes. Altogether the differences between the different methods were of small magnitude and mostly non-significant. We therefore conclude that the two methods produce estimates that are virtually equivalent and that personality measurement is less susceptible to mode effects than other concepts, e.g. alcohol and drug use (Aquilino, 1994) or body weight (Kroh, 2005).

References

Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly*, *58*(2), 210. doi:10.1086/269419

Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. doi:10.1080/10705510903008204

Gerlitz, J.-Y. & Schupp, J. (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Berlin.

Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory (BFI-S). *Journal of Research in Personality*, *46*(3), 355–359. doi:10.1016/j.jrp.2012.03.008

Heineck, G. (2011). Does it pay to be nice? Personality and earnings in the United Kingdom. *Industrial and Labor Relations Review*, *64*(5), 1020–1038. Retrieved from http://www.jstor.org/stable/41343711

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. doi:10.1007/BF02289447

Kroh, M. (2005). Intervieweffekte bei der Erhebung des Körpergewichts in Bevölkerungsumfragen. *Das Gesundheitswesen*, *67*(08/09), 646–655. doi:10.1055/s-2005-858601

Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior Research Methods*, *43*(2), 548–567. doi:10.3758/s13428-011-0066-z

Leeuw, E. D. d. (1992). *Data quality in mail, telephone and face to face surveys.* Amsterdam: TT-Publikaties.

McCrae, R. R. & Costa, P. T. J. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*(5), 509–516.

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah N.J.: L. Erlbaum Associates.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67–88). Lawrence Erlbaum.

Richter, D. & Schupp, J. (2012). *SOEP Innovation Sample (SOEP-IS): Description, Structure and Documentation.* SOEPpapers 463, DIW Berlin.

Sawhney, G. & Cigularov, K. P. (2014). Measurement equivalence and latent mean differences of personality scores across different media and proctoring administration conditions. *Computers in Human Behavior*, *36*, 412–421.

Schaeffer, N., Dykema, J., & Maynard, D. (2010). Interviewers and interviewing. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (pp. 437–470). Binley, UK: Emerald.

Uysal, S. D. & Pohlmeier, W. (2011). Unemployment duration and personality. *Journal of Economic Psychology*, *32*(6), 980–992.