The German Socio-Economic Panel study

888 2016

SOEPpapers

on Multidisciplinary Panel Data Research

SOEP — The German Socio-Economic Panel study at DIW Berlin 888-2016

Response Error in a Web Survey and a Mailed Questionnaire: The Role of Cognitive Functioning

Martin Kroh, Denise Lüdtke, Sandra Düzel, Florin Winter

DIW SOEP

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: soeppapers@diw.de

Response Error in a Web Survey and a Mailed Questionnaire:

The Role of Cognitive Functioning

Martin Kroh[1], Denise Lüdtke[2], Sandra Düzel[3], and Florin Winter[4]

1 Prof. Dr. Martin Kroh, Deutsches Institut für Wirtschaftsforschung (DIW Berlin) and Humboldt-Universität zu Berlin, Mohrenstraße 58, 10117 Berlin, mkroh@diw.de

2 Dr. Denise Lüdtke, Deutsches Institut für Wirtschaftsforschung (DIW Berlin), Mohrenstraße 58, 10117 Berlin

3 Dr. Sandra Düzel, Max-Planck-Institut für Bildungsforschung, Lentzeallee 94 14195 Berlin, duezel@mpib-berlin.mpg.de

4 Florin Winter, Deutsches Institut für Wirtschaftsforschung (DIW Berlin) and Humboldt-Universität zu Berlin, Mohrenstraße 58, 10117 Berlin

This study was supported by the German Federal Ministry of Education and Research (http://www.bmbf.de/en/index.php) under grant number 16SV5537. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Web-based interviewing is gradually replacing traditional modes of data collection, in particular telephone and mailed surveys. This global trend takes place despite the fact that established knowledge of its consequences on response error is incomplete. This paper studies differences between a web (CAWI) and a mailed version (MAIL) of a questionnaire in various forms of response error, namely item nonresponse, satisficing, person-reliability, and social desirable responding. We posit 1) that response error depends on respondents cognitive functioning, namely in the domains of global reading abilities, fluid intelligence, as well as working and episodic memory; and 2) that these effects differ across modes of data collection with generally higher prevalence in the CAWI mode since this mode is more demanding.

The analysis builds on a randomized mode experiment implemented in the context of the Berlin Aging Study II (BASE-II), a survey that primarily focuses on multidimensional processes of physical and mental aging (see Bertram et al. 2014). The analysis reveals a high impact of cognitive functioning at the various stages of the survey response process. While we do found moderate mode-differences in response error, such as higher item nonresponse rates in the CAWI mode, we did not find cognitive functioning to be a better predictor of response error in web-based interviewing.

Keywords: Mixed-mode design, CAWI, cognitive functioning, response quality

1. Motivation

Ongoing technological advancement is reflected in the dominant modes of data collection in surveys. In face-to-face settings, computer-assisted personal interviewing (CAPI) is gradually replacing paper-and-pencil personal interviewing (PAPI), while in self-administered interviews, web-based formats (CAWI) are increasingly replacing mailed questionnaires. In many countries, web-based interviewing has become a critically important mode of data collection in social and market research in only 10 years' time (see Table 1 for the German case).

Table 1 about here

The main advantage of computer assisted web interviewing (CAWI), compared to other modes of data collection, is time and cost efficiency (Dillman 2000; Tourangeau et al. 2013). CAWI proponents also cite audio-visual options (Tourangeau 2004) and the ability to contact both remote populations and specific subpopulations as advantages (Wright 2005). Drawbacks are usually associated with selective coverage of the target population, namely the exclusion of persons without internet access (Couper 2001; Couper et al. 2007) – a problem that loses relevance with comprehensive internet penetration. Moreover, non-response tends to be higher in CAWI mode (Fricker and Schonlau 2002; Shih and Fan 2008; Fan and Yan 2010). Hence, within the conceptual framework of "total survey error," much attention is devoted to the question of selective representation due to CAWI mode (coverage, sampling, non-response error). However, relatively little is known about how CAWI affects response error.

With the purpose of analysing this CAWI mode effect, we decided to compare CAWI with another self-administered questionnaire. Although the comparison with the face-to-face mode seems useful given that this mode still represents the mainly used mode in surveys, it nevertheless differs strongly from self-administered questionnaires both in terms of costs and interviewer assistance. Therefore we decided to compare the CAWI mode with the MAIL mode as a likewise self-administered questionnaire that differs strongly from CAWI with respect to the extent of technology used.

Previous studies on mode effects primarily focus on self-administered versus interviewer administered questionnaires in order to reveal mode effects regarding item non-response, social desirability, and response styles typical for satisficing (Boyer et al. 2002; Kiesler and Sproull 1986; Klassen and Jacobs 2001; Kwak and Radler 2002; de Leeuw 1992). However, comparisons of web surveys with mailed questionnaires regarding mode effects are scarce. We intend to fill this research gap with the study at hand. We implemented a randomized mode experiment during the 2012 BASE-II Study that facilitates the disentanglement of the differences between CAWI and MAIL regarding response error and analysing the role of cognitive abilities on these differences. For this purpose, we identify several forms of response error – specifically item nonresponse, social desirability, response styles, and low person-reliability – and estimate the mode effects in these response errors.

Taking into account that different modes require different levels of cognitive skills, the inclusion of cognitive functioning indicators should facilitate better understanding of the underlying factors for mode differences in data quality. The comprehensive assessment of cognitive functioning in BASE-II enables us to study cognitive effects on response performance in much more detailed fashion than studies previously conducted in this field. For our study, we selected several measures of fluid intelligence – visuo-spatial working memory, episodic memory, and global reading abilities – as each are unique abilities contributing to specific steps in the survey response process. Additionally, these measures can also be linked to the specific requirements related to survey participation in CAWI. In addition to cognitive abilities, we use measures of motivation in our analyses since the satisficing theory postulates that it is not only cognitive limitations, but also a lack of motivation in the respondent that might lead to response error (Krosnick 1991). Moreover, it is well know from psychology research that motivational states are important indicators facilitating the attainment of consequential goals (Wrosch and Scheier 2003), e.g. to maintain attention and to control responses during the self-administered interview.

2. State of Research

For the purpose of giving an adequate answer to survey questions, respondents must successfully integrate multiple sources of information, which in turn requires different dimensions of cognitive functioning. If respondents are not able or not willing to engage in these cognitive processes necessary for survey responding, survey reports are at risk of being plagued by response error (Krosnick 1991, Biemer and Lyberg 2003). Instead of engaging thoughtfully in the answering process, respondents with low cognitive skills or low motivation are likely to use the strategy of satisficing, which means to "simply provide a satisfactory answer instead" (Krosnick 1991, 213). Satisficing "can take the form of either (1) incomplete or biased information retrieval and/or information integration, or (2) no information retrieval or integration at all" (Krosnick 1991, 213). Cognitive demands required in the survey response process depend in part on the difficulty of the task, which in turn partly depends upon the survey mode (Roberts 2007). In section 2.1 we report previous findings regarding the effect of cognitive functioning on response error. Section 2.2 reports on the relationship between survey mode and response error, while section 2.3 combines both research strands and discusses possible interaction effects between survey mode and cognitive functioning on response error.

2.1 Response Error and Cognitive Functioning

Since comprehensive measures of cognitive functioning are scarce in large-scale surveys, studies on response error mostly rely on proxy information for cognitive abilities. However, two studies use test scores on cognitive skills: Fuchs (2009) studies the impact of cognitive functioning on item nonresponse using elaborated measures of cognitive skills from the first wave of the Berlin Aging Study (BASE). He confirms that lowered cognitive abilities are related to higher non response rates. Lechner and Rammstedt (2015) reveal a negative association between cognitive ability (measured with the Digit Symbol Substitution Test) and acquiescence in survey responses as of elderly respondents.

The BASE-II study provides a multitude of measures on cognitive functioning enabling us to elaborate on the impact of specific cognitive skills on response error at different stages in the survey answering process. Tourangeau, et al. (2000, 8) identified four cognitive processes that have to be performed in order to answer to survey questions adequately (see Table 2). The first process is the *comprehension* of the questions, which includes attending to questions and instructions, representing the logical form of questions, identifying the question focus and linking key terms to the relevant concepts.

This first process requires text comprehension and global reading skills. Overall global *reading ability* was assessed in BASE-II using a speeded sentence reading test (Bergmann and Wimmer 2008). A central part of the test is the decision whether presented sentences are wrong or right.

The number of correctly answered sentences within three minutes gives evidence for a causal link between reading speed and reading comprehension. Thus, rapid, accurate readers may expend less of their limited cognitive resources (e.g., attention, working memory) trying to identify words (i.e., fewer resources spent decoding, blending, segmenting, or using context cues to identify the word) than slower readers. As a result, rapid and accurate readers may have more cognitive resources available to apply to comprehension. Additionally, information remains in working memory for a limited amount of time. Thus, individuals who read less rapidly may have more difficulty synthesizing previously read material with material read later, because the previously read material will be less accessible (e.g., Breznitz 1987; Daneman and Carpenter 1980; LaBerge and Samuels 1974; Rasinski 2000; Skinner et al. 1998; Stanovich 1986).

The second process, *retrieval* of relevant information, includes generating a retrieval strategy, retrieving specific, generic memories and filling in missing details. For this process, specific *memory functions* such as visuo-spatial working memory and processing speed are necessary to mentally navigate through the questionnaire, maintain the relevant question and update new information as needed. Working Memory refers to the temporary maintenance and manipulation of information (Baddeley, 2012) in order to reach a specific goal. It involves processes such as to ignore irrelevant information (distractors) in order to maintain behaviour that focuses and serves only the relevant task.

Processing speed is measured by the rate at which tasks can be performed and relates to the ability to process information automatically and, therefore, speedily. There is consensus in cognitive aging literature that performance on many information-processing tasks is slowed in old age (Salthouse 1985). It is suggested that this phenomenon is caused by a general decrease in processing rate with age, but this slowing is less pronounced in tasks requiring lexical decisions (Myerson et al. 1992).

The third process is *judgment*, which means to remember and recall e.g. autobiographic memories or past experiences, in order to make required judgments and decisions in response to a question. Central to this process is the assessment of completeness and relevance of memories, the drawing of inferences based on accessibility, the integration of material retrieved and the making of an estimate based on partial retrieval. These processes are critically dependent on *Episodic Memory* functioning (Tulving 2002), as Episodic Memory can be described as the long-term storage of events that are bound to particular times and places in the past (Tulving 1972).

The fourth process can be described as the selection and *reporting* of an answer, thus including the mapping of a judgment onto response categories and the editing of a response. This process requires a certain capacity to think logically and solve problems in novel situations, independent of acquired knowledge, which is called fluid intelligence (Cattel 1971; Lindenberger et al. 1993).

Table 2 about here

2.2 Survey Mode and Response Error

Previous studies of mode effects on response error regarding mode differences between CAWI and MAIL focus on item nonresponse, while few to no studies discuss scale reliability, social desirability, and response styles.

Regarding item completion rates in comparison between web mode and mailed mode, results are ambiguous. A number of studies report less item non-response in the web mode (Boyer et al.

2002; Kiesler and Sproull 1986; Klassen and Jacobs 2001; Kwak and Radler 2002; Schaefer and Dillman 1998; Barrios et al. 2010). However, newer studies show no differences in item non-response between the two modes (Börkan 2010; Lin and Ryzin 2012). Findings by Miller and Dillman (2012) indicate that mode effects regarding item nonresponse might depend on the content of the questions.

Results for scale reliability are also ambivalent, showing higher pairwise correlations between items either in the MAIL mode (Cole 2005; Lin and Ryzin 2012) or in the web based mode (Börkan 2010). Still others detect no systematic differences between modes (Fouladi et al. 2002, 208).

Social desirable responding is in general lower in self-administered questionnaires. In conducting a meta-analysis, Tourangeau et al. (2013, 142) find a positive, but only very small effect of web questionnaires compared to paper questionnaires regarding the revelation of sensitive information.

In sum, studies of mode effects on response error have inconsistent results. These inconsistencies might be traced back to different survey and question topics and other design features that seem to interact with the survey mode. Moreover, the different samples investigated, implying different levels of cognitive functioning, might also be the source of inconsistent results.

2.3 Cognitive Ability and Mode Effects

As already argued, the process of answering survey questions requires specific cognitive skills. However, these cognitive requirements depend in part on survey mode (Roberts 2007). Although the largest differences in cognitive requirements exist between interviewer-administered questionnaires and self-administered questionnaires, since the former present the questions orally, also two self-administered modes CAWI and MAIL may vary in cognitive requirements (compare Groves et al. 2009). Foremost, CAWI respondents need to have good technology skills, as well as internet, and computer skills. An overview of studies regarding the relationship between cognitive skills and computer task performance is provided by Czaja (1997). She reviews, in particular, the measures of fluid intelligence that are related to computer proficiency. Namely, two measures of fluid intelligence could be related to computer proficiency: Space and Letter Series subscales. Moreover, spatial working memory indicators are associated with computer proficiency, for instance with text-editing. The ability to temporarily maintain relevant information in mind is crucial for the performance of many everyday tasks but also for providing reliable answers to a web questionnaire.

Regarding internet proficiency, Morris (2007) considered disabilities of either physical or mental nature as causes for perceived barriers of internet access. For instance, they argue that low memory or concentration skills could function as a barrier for internet access.

Further, there is substantial evidence that there are age-related declines in most component processes of cognition including: attentional processes, working memory, information processing speed, encoding and retrieval processes in memory, and text comprehension (e.g. Grady 2008; Mather 2015). Since computer tasks are primarily characterized by their mental demands, decrements in these component processes could place older adults at a disadvantage with respect to performing computer-interactive tasks.

Overall, the technological skills associated with CAWI mode, such as working memory and fluid intelligence, overlap with coginitive functions that are also relevant for the survey response process in general. Hence, we expect that individual differences in cognitve ability are better predictors of response error in CAWI mode than in MAIL mode.

3. Data

The analysis draws on survey data emanating from the Berlin Aging Study II (BASE-II). This longitudinal survey focuses primarily on multidimensional processes of physical and mental aging (see Bertram et al. 2014). The BASE-II sample is a non-probability sample consisting of 1600 elderly (age 60 to 80) and a reference group of 600 younger residents (age 20 to 35) living in greater Berlin. To compensate potential processes of selectivity in BASE-II, we compared the data with a large, representative reference study, the German Socio-Economic Panel (SOEP, see Wagner et al. 2007)), and analysed differences in characteristics of participants in BASE-II compared to those from the SOEP. Based on this selectivity analysis, we then generated propensity score weights that adjust for the selectivity in the BASE-II survey. In addition, we adjusted the weights of the BASE-II sample to statistical information from the Federal Statistical Office so that the BASE-II study has the same totals as the official statistics. These procedures ensure generalizability of our findings due to representativeness of the BASE-II sample for the German population regarding common socio-demographic characteristics as well as BASE-II specific factors, namely health-related factors, personality facets, attitudes, and neighbourhood characteristics. A detailed description of the weighting procedure is provided in Saßenroth et al. (2013).

Respondents repeatedly participate in medical check-ups as well as mental and motoric testing at centralized test sites. The cognitive testing of the participants was carried out by the Max Planck Institute for Human Research located in Berlin between 2012 and 2014. Participants were invited to two sessions, within an exact interval of 7 days at the same session time to avoid circadian confounding effects on performance. Each session lasted about 3.5 hours, during which participants were tested in groups of about six individuals. The cognitive battery covers a wide range of functional domains such as attention and processing speed, reading, learning and episodic memory, working memory, as well as decision making.

To collect information about participants' economic position, social relations, and biography, participants and all members of their household were asked to fill in an individual questionnaire and a household questionnaire. The data collection of the socio-economic part of BASE-II was carried out by TNS Infratest for DIW Berlin. Participants received questionnaires up to 5 times between 2008 and 2014 (see Böckenhoff et al. 2013).

Since the socio-economic data collection in 2012 included an experimental design to study mode effects, we focus on this wave in our analyses. Moreover, we exclude all individuals who failed to participate in either the socio-economic part in 2012 or the psychological testing from our analysis. This decisions results in a sample of N= 1,398 persons.

4. Experimental Design

For the field experiment in 2012, we randomly assigned all households to either the CAWI or the MAIL mode. As the randomization of assignment operated at the head of the household level, all corresponding household members were allocated to the same group. The assignment was not mandatory, as we used incentives to motivate the respondents to choose the survey mode corresponding to their assignment. Although a complete assignment by offering only a single mode option in each group was, of course, feasible, it might have reduced survey participation rates. The survey was not primarily designed for methodological research, but rather to investigate processes of ageing.

Instead of the usually paid 10 Euros, we offered 15 Euros for an individual questionnaire filled-in according to the assigned mode. The randomized assignment was stratified by the prior survey mode in order to avoid confounding effects. In the two previous waves, 2008 and 2009, mixed-modes were used. Table 3 reports that 110 persons in our sample were interviewed face-to-face in 2009 (PAPI plus CAPI), the majority of 829 persons are former MAIL and CAWI respondents, 120 temporarily refused to participate in the 2009 survey, and 339 are newly recruited participants in 2012.[1]

Table 3 about here

Table 4 displays the choice of survey mode in 2012 by assignment group. Our experimental design that provides different monetary incentives for respondents in the CAWI and MAIL mode produces levels of non-compliance to the allocation rule that are remarkable given the weak incentive structure. In the CAWI assignment, 244 persons (36 percent) used the MAIL mode, while approximately 441 persons (64 percent) used the CAWI mode. The incentive worked even better for the MAIL assignment, as about 611 persons (86 percent) from the MAIL group, who participated in the survey, filled-in the questionnaire using MAIL mode. The marginal effect of

[1] The initial advance letter already provided an access code for the CAWI interview and respondents could immediately start with the interview after having received it. In contrast, respondents who decided to answer the questionnaire in MAIL mode had to inform TNS Infratest about their decision. In a second step, TNS Infratest then sent MAIL questionnaires to the household.

the randomized incentives of roughly 50 percentage points corresponds to an increase in the odds ratio of mode choice by a factor of 10.8. Refusal rates were comparable across assignments, with 8 percent for the MAIL mode assignment and 9 percent for the CAWI mode assignment.

Table 4 about here

Table 5 displays the compliance rates to the mode assignment in 2012 by survey mode in 2009. It can be seen that the compliance with the mode assignment in 2012 differs with regard to the survey mode chosen in 2009. For instance, from those respondents assigned to the MAIL mode, 99 percent who filled in the questionnaire in 2009 in MAIL mode complied with the assignment, while only 78 percent who formerly filled-in the questionnaire in CAWI complied with the MAIL assignment. More obvious are the differences in the CAWI assignment. Only 35 percent of respondents who selected MAIL mode in 2009 and were assigned to the CAWI group in 2012 indeed filled in their questionnaire online, whereas 85 percent of the respondents from the former CAWI mode complied with the CAWI assignment in 2012. These differences in compliance rates — as one would expect — are not statistically independent from survey mode used in 2009 (for MAIL assignment: $Chi^2(5) = 60.0$, $p < 0.001$; for CAWI assignment: $Chi2(5) = 138.0$, $p < 0.001$). The identification of the causal mode effect in the following section therefore not only considers non-compliance by an instrumental variables approach, but also selective compliance rates by mode choice in the previous wave.

Table 5 about here

5. Analysis

We investigate the impact of mode choice on the quality of respondents' answering behaviour both by OLS regressions and by an instrumental variables approach. We focus on item non-response, social desirability, response styles, and person-reliability. We selected specific measurements of the participants' cognitive abilities to calculate their effect on indicators of data quality. To investigate whether or not cognitive ability is more relevant in the CAWI mode, we include interactions between survey mode and cognitive abilities in our models.

5.1 Indicators for data quality

We aim at examining data quality from several perspectives, thus calculating different indicators that inform about different aspects of data quality. We chose the indicators in accordance with the theory on satisficing (Krosnick 1991; Krosnick et al. 1996). Satisficing behaviour can result in item non-response, mental coin flipping, and in different response styles, such as a preference for the middle response category and extremeness (see Roberts 2007 for an overview). Thus, we look at the item nonresponse rate, response styles, and the person-reliability. Moreover, we use social desirability as another indicator for poor answer quality (compare Paulhus 1984). We built our indicators as follows:

Item Nonresponse

We count all missing values for all numerical items in the survey. The item non-response rate is then calculated as proportion between the missing values and the total number of numerical items in the survey (minus the items that didn't apply to the respondent). The actual number of items that applied to the respondents ranged from 236 to 346 with an item non-response rate ranging from 0 to .37 and a mean of .03. A higher item nonresponse rate indicates poorer answer quality.

Social Desirability

Low motivation can not only result in satisficing, but also in social desirability, if respondents are not willing to disclose their honest answers to specific questions (Sudman and Bradburn 1974). Socially desirable responding is "the tendency to give answers that make the respondent look good" (Paulhus 1991, 17). A short inventory based on the 40-item "Balanced Inventory of Desirable Responding" (BIDR) by Paulhus (1991) was used to detect the tendency for social desirable responding. We build an averaged additive scale. Values range from 1 to 7 with high scores indicating a strong tendency for social desirability and, thus, poor answer quality. Although social desirability items are expected to produce lots of missings, in our data 98 percent of the respondents answered all four items. The mean scale score amounts to 4.73 with a standard deviation of 1.11.

Response Styles: Midpoint and Extremeness.

While normal responding would imply that respondents use the full scope of the scales presented, response styles are present if monotonic response behaviour is shown. Response style can be defined as the tendency to respond to questions on the basis other than their content and doing so consistently across time and situations (Paulhus 1991). Response styles lead to systematic errors in measurement, with different effects for different response styles (van Vaerenbergh and Thomas 2013). We consider in our analysis two types of response styles, namely 1) the preference for the middle response category, and 2) extremeness. Extremeness can be described as “a preference for selecting answers from the end points of a scale” (Roberts 2007, 13). As basis for calculating the two response styles we used all items in the survey that rely on rating scales (that is 141 items). We count the number of extreme or midpoint answers for each respondent over all 141 items. The indicator for extremeness ranges from .10 to .84 with a mean of .44. The indicator for midpoint responses ranges from .02 to .51 with a mean of .20. High scores on both indicators imply low data quality.

Person-Reliability

One type of satisficing behaviour is mental coin-flipping which means that respondents do not select the response categories by thoughtful consideration but randomly (Krosnick 1991, 220). Mental coin-flipping should result in a low person-reliability (see Krosnick et al. 1996). We use statistics of the internal consistency of multi-item constructs, referred to as person-fit statistics, to evaluate reliability of respondents’ answers. The underlying objective of person-fit statistics is to identify aberrant response patterns in multi-item measures (like the Big 5 trait of extraversion); that is, response patterns that deviate from the estimated item response model based on Item Response Theory (IRT) (Van Vaerenbergh and Thomas 2013). The parametric person-fit statistics used throughout this paper are used to establish the difference between the observed and the expected item scores over a number of items (for binary response options, see Meijer and Sijtsma (2001) and Karabatsos (2003), for polytomous items see van Krimpen-Stoop and Meijer (2002); Dagohoy (2005); Conijn (2013)). This indicator helps to detect unexpected answers (e.g. agreeing with both the regular and the reverse coded item within one multi-item construct), which might be the result of mental coin flipping.

For this study, a total of 20 multi-item scales could be obtained. The constructs range from 2 to 8 items, comprising attitude and personality questions, e.g. the Big 5 personality traits, life satisfaction, subjective health (mentally and physically), and optimism. The calculation of the l-person-fit statistic plus the multi-item constructs used is based in the paper by Kroh, Winter and Schupp (2016). The indicator for person-reliability ranges from -1.91 to .90 with a mean of .09. High values of the indicator reflect a high person-reliability.

5.2 Indicators of Cognitive Functioning

The cognitive tasks were assessed on two sessions, within an exact interval of 7 days at the same time of day to avoid circadian confounding effects on performance. Each session lasted for about 3.5 hours. Participants were tested in groups of 4–6 individuals. The cognitive battery included measures of learning and memory performance, attention/processing speed, working memory, executive functioning, and perceptual speed. The group was instructed through a standardized session protocol. Each task started with a practice trial to make sure that all participants understood the task. Responses were given via button boxes, mouse, or keyboard. For the purpose of the present study, only those cognitive tasks included in the study are described in detail below (see also: Düzel et al 2015; Schmiedek et al. 2010; Lindenberger et al. 1993).

Overall Reading Ability

We used the overall reading ability as measure for cognitive skills required for component 1 in the survey response process (comprehension, see Table 2). The overall reading ability was assessed using a speeded sentence reading test (Bergmann and Wimmer 2008) for which norm data is currently evaluated and planned to be published as the adult version of the Salzburger

Lese-Screenings (Auer et al. 2004; Mayringer and Wimmer 2002). The sentences are of relatively easy content (e.g., Ein Nashorn ist ein Blechblasinstrument [A rhinoceros is a brass instrument]) but become more complex in their structure as the test proceeds. Participants were instructed to evaluate sentences as quickly and accurately as possible and to give a yes-response when the content made sense and a no-response if not. The dependent variable was the number of correctly answered sentences within three minutes.

Working Memory

As measure for cognitive skills necessary to achieve the second component of the survey response process (Retrieval; see Table 2), we selected two working memory tasks that measure spatial updating performance and processing speed (measured with the digit symbol substitution test). In each block of the Spatial Updating Task, a display of two (respectively three) 3x3 grids is shown for 4 seconds in each of which one blue dot was present in one of the nine locations. Those two (or three) locations have to be memorized and updated according to shifting operations, which were indicated by arrows appearing below the corresponding field. PT of the arrows is 2.5seconds with an ISI of 0.5seconds. After six updating operations, the two (or three) grids reappeared and the resulting end positions had to be clicked on. After 10 practice blocks with memory load two (or three) grids, ten test blocks with load two (or three) were conducted and used for scoring. The average of percent correct placements was used as the working memory score in this study (see also Schmiedek et al. 2010).

Additionally, the Digit Symbol Substitution Test (WAIS, Wechsler 1981) is a neuropsychological test sensitive to brain damage, dementia, age and depression. It consists of (e.g. nine) digit-symbol pairs followed by a list of digits. Under each digit the subjects write down the corresponding symbol as fast as possible. The number of correct symbols within the allowed time (90 seconds) is measured. The test sheet was enlarged by 100 percent to reduce perceptual and motor problems.

Episodic Memory

For assessing the third component of the survey response process (Judgment; see Table 2), we used a latent factor score on episodic memory performance (by applying a structural equation model; SEM). The factor consists of four cognitive tasks (Scene Encoding Task, Verbal Learning and Memory Test, Face-Profession Task and Object Location Memory Task). Within the Scene Encoding Task, participants performed an incidental encoding task with eighty-eight complex, grey-scaled, images (44 indoor and 44 outdoor scenes; mean grey value 127, S.D. 75) of neutral emotional valence sequentially presented each for 2.5s. Participants performed an indoor/outdoor judgment to each image by button presses (left index finger for inside scenes and right index finger for outside scenes). In the retrieval tests, 44 old images (22 indoor-22 outdoor) presented at encoding task were randomly presented together with 44 novel distractor images (22 indoor-22 outdoor) for 2.5s. After each presentation participants rated their confidence of recognition memory on a scale ranging from 1 to 5 ("1": sure new; "2": may be new; "3": unsure whether old or new; "4": old; "5": sure old). There were no time restrictions for confidence ratings. Recognition memory performance for scenes (hit minus false alarms) was assessed after a delay of approximately 2.5 hours (short-delay memory), and used as manifest variable in the SEM. The Verbal Learning and Memory Test assesses auditory verbal learning of 15 words including five learning trials, early recall, an interference list after five learning trials, free recall tests directly after an interference list as well as 30 min later (late recall), and a final recognition test [3]. The sum of items recalled across trial 1 to 5 provides a measure of overall learning performance, which was used here as a manifest variable. The face-profession task assesses associative binding on the basis of recognition of incidental encoded face-profession pairs. During the study phase 45 face-profession pairs were each presented for 3.5s on the computer screen and the participants had to indicate via button presses whether the faces matched to the profession or not. After a 3 minute delay between study and test phase 54 face-profession pairs were presented consisting of 27 old pairs, 9 new pairs, and 18 new-arranged pairs (here the same face is shown, but associated with a new profession). The participants were asked to decide whether a given face-profession combination has been seen before or not and to rate the confidence of their decision on a three point scale ranging from 1 = not sure to 3 = very sure. Recognition memory for the rearranged face-profession pairs (hit minus false alarms) served as manifest variable in the model. In the Object Location Memory Task sequences of 12 coloured photographs of real-world objects were displayed at different locations in a 6 by 6 grid. After presentation, objects appeared at the side of the screen and had to be moved to the correct locations by clicking on the objects and the locations with the computer mouse. One practice trial and two test trials were included. The sum of correct placements across the two test trials is used as manifest variable (see also Düzel et al. 2015, supplemental material).

Fluid Intelligence

In order to assess underlying cognitive abilities of component 4 of the survey response process (Response, see Table 2), we selected a global fluid intelligence marker and established a latent factor score for fluid intelligence performance by using SEM. For this purpose, scores from the Practical Problems Task, the Figural Analogies test, and the Letter Series Task were selected. The Practical Problems Task consists of 22 items depicting everyday problems such as the hours of a bus schedule, instructions for medication, a warranty for a technical appliance, a train map, as well as other forms and tables. For each item, the problems were presented in the upper part of the screen, and five response alternatives were shown in the lower part. Subjects responded by clicking with the mouse-arrow on one of the five alternatives. A single practice item was provided. The test phase was terminated when subjects made three consecutive errors, or when they reached the maximum time limit of 10 minutes, or after they had answered the last item of the test. Items are ordered by difficulty (see also Lindenberger et al. 1993). The 22 items in the Figural Analogies test followed the format "A is to B as C is to ?". One figure-pair was presented in the upper left part of the screen and an incomplete figure was shown beside. The participants had to apply the same rule that was applied to the complete figure-pair by choosing one of the five alternative responses which were presented in the lower part. Subjects entered their response by clicking with the mouse-arrow on one of the five alternatives. Before the test phase, instructions and three practice items were given. The test phase was terminated when subjects made three consecutive false responses, when they reached the maximum time limit (10 min), or after they had answered the last item of the test. Items are ordered by difficulty (see also Lindenberger et al. 1993). The Letter Series Task consisted of 22 items. Each item contained five letters followed by a question mark (e.g., c e g i k ?). Items were displayed in the upper half of the screen, and five response alternatives were presented in the lower half. Items followed simple rules such as +1, -1, +2, or +2 +1. Subjects entered their response by touching one of the five answer alternatives. The score was based on the total number of correct responses. Before the test phase, instructions and three practice items were given. The test phase was terminated when subjects made three consecutive false responses, when they reached the maximum time limit (6 min), or after they had answered the last item of the test. Items are ordered by difficulty. Sample items were used with respect to tests related to speed, reasoning, and knowledge (see also Lindenberger et al. 1993).

5.3 Control Variables

We include the level of education in addition to the sophisticated measures of cognitive functioning in our models as well as gender and age of respondents. Indeed, studies on surveys among the elderly found that older respondents are more susceptible to social desirability tendencies (Campbell et al. 1976, 110), non-opinion answers (Schuman and Presser 1981), item nonresponse (Colsher and Wallace 1989), and “yea-saying” tendencies (Ross and Mirowsky 1984).

We furthermore control for the survey mode in 2009, as the choice of the CAWI mode in 2009 might indicate computer literacy. Moreover, we already showed that compliance with the mode assignment is not statistically independent from mode choice in 2009 (see Table 5). The self-reported interview duration may reflect the degree of respondents’ motivation to answer to the questions thoughtfully. For instance Barthelt and Bauknecht (2011) found some relationship between speeding in interviews and satisficing. Since measures of cognitive functioning were collected over a 21 month period, we also control for the time difference between questionnaire completion and cognitive testing.

A good performance in survey responding not only depends on respondents’ cognitive abilities, but also on their motivation to maintain 'time effective' attention to one item but also be able to carry on to the next item / task. Thus, we also control for those kinds of motivational states of the respondents by including measures of goal engagement and goal disengagement (Schulz and Heckhausen 1996). On a short term level, it is necessary for the individual to increase his or her effort to achieve the selected goal (= goal engagement). Once the deadline is passed and the goal (e.g. response to the item) has not been achieved, disengagement from the old goal and selection of a new one is required (= goal disengagement). Otherwise, valuable time, energy, and motivational resources are expended on unattainable goals (Schulz and Heckhausen 1996). Therefore, we expect both, goal engagement and goal disengagement, to have positive effects on survey responding.

Additionally, we decided to control for the dimension of conscientiousness, part of the Big 5 inventory. Note that we excluded the dimension of conscientiousness in our dependent variable of person-reliability to avoid artificial correlations. We assume that persons with a high score on the trait of conscientiousness should generally show less satisficing and higher levels of person-reliability since respondents' motivation to engage in the response process thoughtfully can be affected by respondents' characteristics as well (Roberts 2007).

The randomized differential incentives are used as exogenous variation in the survey mode. That means they represent the instruments in our instrumental variable regressions.

6. Results

Table 6 displays the differences in respondents' characteristics between CAWI and MAIL group. The table supports common views and findings: CAWI respondents are more likely male and highly educated. MAIL respondents are on average 9 years older than CAWI respondents. Regarding cognitive functioning, on average CAWI respondents achieve better test scores in all tests considered. However, CAWI respondents are less conscientious compared to MAIL respondents. Regarding motivation, CAWI respondents have lower scores on goal engagement but also lower scores on goal disengagement compared to MAIL respondents. This means that persons in the CAWI mode have difficulties dealing with new issues and with dropping old issues.

Table 6 about here

In the bivariate setting, data quality differs significantly by the choice of survey mode in 2012. Table 7a displays that the CAWI mode is significantly positive related to extremeness in answers. In contrast, social desirability and the prevalence for middle answer categories are significantly negatively associated with responses in the CAWI mode. The finding regarding social desirability confirms Tourangeau et al. (2013, 142), who find a very small positive effect of web questionnaires compared to mailed questionnaires regarding the revelation of sensitive information. The item nonresponse rate does not differ significantly between modes, which is in line with the studies by Börkan (2010) and by Lin and Ryzin (2012). Person-reliability does not differ significantly between the modes either, as also found by Fouladi et al. (2002).

Tables 7a and 7b about here

Table 7b displays the different aspects of response error by cognitive abilities. For the purpose of better readability, we collapsed each of the five cognitive test scores in a 0/1-variable indicating whether the ability was above average (=1) or not (=0). The mean scores of three from the five data quality indicators vary by cognitive abilities. If cognitive abilities are above average, the tendency for social desirability as well as the tendency to choose the midpoint category is weaker; the item nonresponse rate is lower. We find no bivariate association between cognitive abilities on the one hand and extremeness and person-reliability on the other hand (see the multivariate analysis for different results).

We employ an instrumental variables approach to estimate a causal mode effect in response error. Respondents chose their survey mode based not just on the selective incentive structure, but also, presumably, on their preferences, Internet access, and computer skills. To prevent these confounding factors from biasing our estimate of mode effects, we instrument the chosen mode of data collection by the randomly allocated incentive scheme. Thus, the first stage of the two-stage IV estimator regresses survey mode on the allocation of the survey mode to decompose survey mode into the part that is determined by the allocation and into the error term, which comprises other criteria for the mode choice (e.g., computer skills). Tests statistics suggest that the strength of the instruments exceeds conventional levels ($F(8, 643) = 35.803$, $p = 0.00$). The reported F statistic refers to the instrumented survey mode. The F statistics of the instrumented interactions between survey mode and cognitive functioning exceed the critical value of 10 as well with $F(8, 643) \geq 25.23$.

In the second stage, we regress the five indicators of response error on the instrumented survey mode of the first-stage regression. Additionally, we control for mode choice in the previous wave, socio-demographics, duration of the interview, conscientiousness, goal engagement vs. disengagement and time difference between cognitive testing and questionnaire answering. To account for conditional effects of cognitive functioning, we also include interactions of mode with cognitive skills.

Tables 8a-d about here

Tables 8a to 8d display the results of the instrumental-variables regression models that estimate the effect of the incentive split with interactions between mode choice in 2012 and different aspects of cognitive functioning. The findings from the bivariate analysis can be affirmed regarding midpoint responses and extremeness as the multivariate analysis reveals that the choice for the CAWI mode turns out to be positively related to extremeness and accordingly negatively related to the tendency to choose midpoint categories. Additionally, the CAWI mode inflates item nonresponse.

Regarding effects of cognitive skills, we find significant positive effects of all five tested cognitive abilities on person-reliability. Thus, the higher the level of cognitive skills, the higher the consistency of answers in multi-item measures. Moreover, all cognitive skills considered in our analyses reduce item nonresponse as well as—with the exception of working memory—the tendency to provide midpoint answers. These findings suggest that the choice of midpoint responses results from problems at the early stages of the response process, namely comprehension and retrieval.

More importantly, however, we find significant interactions between cognitive skills and survey mode in single cases (Tables 8a through 8d). While in most cases these interaction terms are insignificant, suggesting that cognitive functioning affects response error rather similarly in CAWI and MAIL mode, we find particularly in models of person reliability significant interaction terms. More specifically, for fluid intelligence and both measures of working memory, namely spatial updating performance and processing speed, we find negative interaction effects with survey mode for person-reliability. While fluid intelligence and working memory increase person reliability, i.e. the consistency of answers, in this MAIL mode, this effect is significantly weaker in the CAWI mode. This finding clearly runs counter our initial expectation that cognitive functioning is more relevant for the response behaviour in web surveys as opposed to mailed questionnaires.

7. Discussion

Motivated by the trend that web-based interviewing is gradually replacing traditional modes of data collection, we investigated the effect of CAWI on data quality. We compared the answer quality in MAIL and CAWI mode using data from the Berlin Aging Study II (BASE-II). We conducted an experiment in which the choice of survey mode was manipulated as we offered monetary incentives that varied in their amount conditional on the choice of data collection mode. Using an instrumental variables approach, we investigated the mode effect on data quality using randomized differential incentives as exogenous variation in the survey mode. We focus on mode effects in item non-response rates, response styles (namely midpoint responding and extremeness), social desirability tendencies, and person-reliability.

Specifically, we conclude that cognitive abilities operating on the first and on the last stage of the survey answering process are influential. That is, global reading abilities managing the comprehension of the questions and instructions, and fluid intelligence operating on the response stage and piloting the mapping of judgments onto response categories.

Contrary to our expectations, we find rather similar effects of cognitive abilities on response error in CAWI and MAIL mode. One notable exception is person-reliability, i.e. the internal consistency of answers. While we do find a positive effect of working memory and fluid intelligence on reliability in MAIL mode, this effect is significantly weaker in CAWI settings.

References

Auer M., G. Gruber, H. Mayringer, and H. Wimmer. 2004. *Salzburger Lese-Screening 5-8. Handbuch*. Manuskript: Salzburg Universität. Available at: http://www.eduhi.at/dl/Salzburger_Lesescreening_Handbuch.pdf

Baddeley, A. 2012. "Working Memory: Theories, Models, and Controversies." *Annual Review of Psychology* 63: 1-29. DOI: 10.1146/annurev-psych-120710-100422.

Barrios, M., A. Villarroya, Á. Borrego, and C. Ollé. 2010. "Response Rates and Data Quality in Web and Mail Surveys Administered to PhD Holders." *Social Science Computer Review* 29 (2): 208-220. DOI: 10.1177/0894439310368031.

Barthelt, S. and J. Bauknecht. 2011. "Effects of Speeding on Satisficing in Mixed-Mode Surveys." 4th Conference of the European Survey Research Association (ESRA). Lausanne, CH.

Bergmann, J. and H. Wimmer. 2008. "A dual-route perspective on poor reading in a regular orthography: Evidence from phonological and orthographic lexical decisions." *Cognitive Neuropsychology* 25(5): 653-676. DOI: 10.1080/02643290802221404.

Bertram, L., A. Böckenhoff, I. Demuth, S. Düzel, R. Eckardt, S.-C. Li, U. Lindenberger, G. Pawalec, T. Siedler, G. G. Wagner, and E. Steinhagen-Thiessen. 2014. "Cohort profile: The Berlin Aging Study II (BASE-II)." *International Journal of Epidemiology* 43 (3): 703-712. DOI: 10.1093/ije/dyt018

Biemer, P. P. and L. E. Lyberg. 2003. *Introduction to Survey Quality*. Hoboken, NJ: Wiley-Interscience.

Böckenhoff, A, D. Saßenroth, M. Kroh, T. Siedler, P. Eibich, and G. G. Wagner. 2013. "The Socio-Economic Module of the Berlin Aging Study II (SOEP-BASE): Description, Structure, and Questionnaire." *SOEPpapers 568*.

Börkan, B. 2010. "The Mode Effect in Mixed-Mode Surveys: Mail and Web Surveys." *Social Science Computer Review* 28 (3): 371-380. DOI: 10.1177/0894439309350698.

Boyer, K. K., J. R. Olson, R. J. Calantone, and E. C. Jackson. 2002. “Print Versus Electronic Surveys: A Comparison of Two Data Collection Methodologies.” *Journal of Operations Management* 20(4): 357-373. DOI: 10.1016/S0272-6963(02)00004-9.

Breznitz, Z. 1987. “Increasing first graders' reading accuracy and comprehension by accelerating their reading rates.” *Journal of Educational Psychology* 79(3): 236-242. DOI: 10.1037/0022-0663.79.3.236.

Campbell, A., P. E. Converse, and W. L. Rogers. 1976. *The Quality of American Life*. New York: Sage.

Cattel, R. 1971. “Estimating modulator indices and state liabilities.” *Multivariate Behavioral Research* 6: 7-33. DOI: 10.1207/s15327906mbr0601_2.

Cole, S. T. 2005. “Comparing Mail and Web-Based Survey Distribution Methods: Results of Surveys to Leisure Travel Retailers.” *Journal of Travel Research* 43: 422-430. DOI: 10.1177/0047287505274655.

Colsher, P. L. and R. B. Wallace. 1989. “Data Quality and Age: Health and Psychobehavioral Correlates of Item Nonresponse and Inconsistent Responses.” *Journal of Gerontology* 44(2): 45-52.

Conijn, J. M. 2013. *Detecting and explaining person misfit in non-cognitive measurement.* Ridderkerk: Ridderprint.

Couper, M. P. 2001. “The Future of Modes of Data Collection.” *Public Opinion Quarterly* 75(5): 889-908. DOI: 10.1093/poq/nfr046.

Couper, M. P., A. Kapteyn, M. Schonlau, and J. Winter. 2007. “Noncoverage and nonresponse in an Internet survey.” *Social Science Research* 36 (1): 131-148. DOI: 10.1016/j.ssresearch.2005.10.002.

Czaja, S. 1997. “Computer Technology and the Older Adult.” In *Handbook of Human Computer Interaction,* edited by M. Helander, T. K. Landauer, and P. Prabhu, 797-812. Amsterdam: North-Holland.

Dagohoy, A. V. T. 2005. *Person Fit for Tests with Polytomous Responses*. Doctoral Thesis. Enschede: Twente University. Available at: http://purl.utwente.nl/publications/51103.

Daneman M. and P. A. Carpenter. 1980. “Individual differences in working memory and reading.” *Journal of Verbal Learning and Verbal Behavior* 19 :450–466. DOI: 10.1016/S0022-5371(80)90312-6.

de Leeuw, E. D. 1992. *Data Quality in Mail, Telephone, and Face to Face Surveys*. Doctoral Thesis. Amsterdam: TT-Publikaties. Available at: http://edithl.home.xs4all.nl/pubs/disseddl.pdf

Dillman, D. A. 2000. *Mail and Internet surveys: The tailored design method*. New York: Wiley. 8317.1985.tb00817.x.

Düzel, S., Voelkle, M. C., Düzel, E., Gerstorf, D., Drewelies, J., Steinhagen-Thiessen, E., Demuth, I., Lindenberger, U. 2015. ”The Subjective Health Horizon Questionnaire (SHH-Q): Assessing Future Time Perspectives for Facets of an Active Lifestyle.” *Gerontology* (in press).

Fan, W. and Z. Yan. 2010. “Factors Affecting Response Rates of the Web Survey: A Systematic Review.” *Computers in Human Behavior* 26 (2): 132-139. DOI:10.1016/j.chb.2009.10.015.

Fouladi, R. T., C. J. Mccarthy, and N. P. Moller. 2002. “Paper-and-Pencil Or Online?: Evaluating Mode Effects on Measures of Emotional Functioning and Attachement.” *Assessment* 9 (2): 204-215.

Fricker, R. and M. Schonlau. 2002. “Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature.” *Field Methods* 14(4): 347–367. DOI: 10.1177/152582202237725.

Fuchs, M. (2009). “Item-Nonresponse in einer Befragung von Alten und Hochbetagten. Der Einfluss von Lebensalter und kognitiven Fähigkeiten.“ *Österreichische Zeitschrift für Soziologie* 9 (special issue): 333-349.

Grady, C. L. 2008. "Cognitive neuroscience of aging." *Annals of the New York Academy of Sciences* 1124: 127-44. DOI: 10.1196/annals.1440.009

Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*. New York: Wiley.

Karabatsos, G. 2003. "Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics." *Applied Measurement in Education* 16(4): 277-298. DOI: 10.1207/S15324818AME1604_2.

Kiesler, S. and L. S. Sproull. 1986. "Response Effects in the Electronic Survey." *Public Opinion Quarterly* 50(3): 402-413. DOI: 10.1086/268992.

Klassen, R. D. and J. Jacobs. 2001. "Experimental Comparison of Web, Electronic and Mail Survey Technologies in Operations Management." *Journal of Operations Management* 19(6): 713-728. DOI: 10.1016/S0272-6963(01)00071-7.

Kroh, M., F. Winter, and J. Schupp. 2016. "*Using Person-Fit Measures to Assess the Impact of Panel Conditioning on Reliability*." Public Opinion Quarterly first published online June 14, 2016 doi:10.1093/poq/nfw025.

Krosnick, J. A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5: 213-236. DOI: 10.1002/acp.2350050305.

Krosnick, J. A., S. Narayan, and W. R. Smith. 1996. "Satisficing in Surveys: Initial Evidence." *New Directions for Evaluation. Special Issue: Advances in Survey Research* 70: 29-44. DOI: 10.1002/ev.1033

Kwak, N. and B. T. Radler. 2002. "A Comparison between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality." *Journal of Official Statistics* 18(2): 257-273.

LaBerge D. and S. J. Samuels. 1974. "Toward a theory of automated information processing in reading." *Cognitive Psychology* 6: 293-323.

Lechner, C. M. and B. Rammstedt. 2015. "Cognitive Ability, Acquiescence, and the Structure of Personality in a Sample of Older Adults." *Psychological Assessment* 27(4): 1301-11. DOI: 10.1037/pas0000151.

Lin, W. and G. van Ryzin. 2012. "Web and Mail Surveys: An Experimental Comparison of Methods for Nonprofit Research." *Nonprofit and Voluntary Sector Quarterly*, 41(6), 1014-1028. DOI: 10.1177/0899764011423840

Lindenberger, U., Mayr, Ul., Kliegl, R. (1993). "Speed and intelligence in old age." *Psychology and Aging* 8 (2): 207-220.

Mather, M. 2015. "The Affective Neuroscience of Aging." *Annual Review of Psychology* 67: 213-238. DOI: 10.1146/annurev-psych-122414-033540.

Mayringer, H. and H. Wimmer. 2002. "No deficits at the point of hemispheric indecision." *Neuropsychologia* 40(7): 701-704.

Meijer, R. R. and K. Sijtsma. 2001. "Methodology Review: Evaluating Person-Fit." *Applied Psychological Measurement 25*(2): 107-135. DOI: 10.1177/01466210122031957.

Millar, M. M. and D. A. Dillman. 2012. "Do Mail and Internet Surveys Produce Different Item Nonresposne Rates? An Experiment Using Random Mode Assignment." *Survey Practice* 5 (2).

Morris, A. 2007. "E-literacy and the grey digital divide: a review with recommendations." *Journal of information literacy* 1 (3): 13 - 28. DOI: http://dx.doi.org/10.11645/1.3.14.

Myerson, J., F. R. Ferraro, S. Hale, and S.D. Lima.1992. "General slowing in semantic priming and word recognition." *Psychology and Aging* 7: 257–270.

Paulhus, D. L. 1984. "Two-component models of socially desirable responding." *Journal of Personality and Social Psychology* 46(3): 598-609. DOI: 10.1037/0022-3514.46.3.598.

Paulhus, D. L. 1991. "Measurement and Control of Response error." In *Measures of personality and social psychological attitudes. Measures of social psychological attitudes*. Edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, 17-59. San Diego, US: Academic Press.

Rasinski, T.V. 2000. “Speed does matter in reading.” *Reading eacher* 54(2): 146-151.

Roberts, C. 2007. “Mixing Modes of Data Collection in Surveys: A Methodological Review.” ESCR National Centre for Research Methods NCRM Methods Review Papers NCRM/008.

Ross, C. E. and J. Mirowsky. 1984. “Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health.” *Journal of Health and Social Behavior* 25(2): 189-197.

Salthouse, T. A. 1985. “Anticipatory processing in transcription typing.” *The Journal of Applied Psychology* 70(2): 264-271.

Saßenroth, D., M. Kroh, and G. G. Wagner. 2013. “Selectivity Processes in and Weights for the Berlin Aging Study II (BASE-II).” *SOEPpapers 608.*

Schaefer, D. and D. Dillman. 1998. “Development of a Standard E-mail Methodology: Results of an Experiment.” *Public Opinion Quarterly* 62: 378-397.

Schmiedek, F., Lövdén, M., Lindenberger, U. 2010. ”Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study.” *Frontiers in Aging Neuroscience*. DOI: 10.3389/fnagi.2010.00027, PMID: 20725526.

Schulz, R. and J. Heckhausen. 1996. “A life span model of successful aging.” *American Psychologist* 51(7):702-14.

Schuman, H. and S. Presser. 1981. *Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context*. San Diego: Academic Press.

Shih, T.-H., and X. Fan. 2008. “Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis.” *Field Methods* 20 (3): 249-271. DOI: 10.1177/1525822X08317085.

Skinner, E. A., M. J. Zimmer-Gembeck, and J.P. Connell. 1998. “Individual differences and the development of perceived control.” *Monographs of the Society for Research in Child Development* 63(2/3): i+iii+v-vi+1-231.

Stanovich, K. E. 1986. “Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy.” *Reading Research Quarterly* 22: 360-407.

Sudman, S. and N. M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine Publishing.

Tourangeau, R. 2004. "Survey Research and Societal Change." *Annual Review of Psychology* 55: 775–801. DOI: 10.1146/annurev.psych.55.090902.142040.

Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.

Tourangeau, R., F. Conrad, and M. P. Couper. 2013. *The Science of Web Surveys*. New York: Oxford University Press.

Tulving, E. 1972. "Episodic and semantic memory". In *Organization of Memory,* edited by E. Tulving and W. Donaldson, 381-402. New York: Academic Press.

Tulving, E. 2002. "Episodic Memory: From Mind to Brain." *Annual Review of Psychology* 53: 1-25. DOI: 10.1146/annurev.psych.53.100901.135114

van Krimpen-Stoop, E. M. L. A. and R. R. Meijer. 2002. "Detection of Person Misfit in Computerized Adaptive Tests with Polytomous Items." *Applied Psychological Measurement* 26(2): 164-180. DOI: 10.1177/01421602026002004.

van Vaerenbergh, Y. and T. D. Thomas. 2013. "Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies." *International Journal of Public Opinion Research*: 1-23. DOI: 10.1093/ijpor/eds021.

Wagner, G. G., J. R. Frick and J. Schupp. 2007. "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements", *Schmollers Jahrbuch*, Vol. 127, pp. 139-169.

Wechsler, D. 1981. *Manual for the Wechsler Adult Intelligence Scale — Revised*. New York: Psychological Corporation.

Wright, K. B. 2005. “Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services.” *Journal of Computer-Mediated Communication* 10 (3). DOI: 10.1111/j.1083-6101.2005.tb00259.x

Wrosch, C. and F. M. Scheier. 2003. “Personality and quality of life: The importance of optimism and goal adjustment” *Quality of Life Research* 12 (Suppl. 1): 59–72.

TABLES

Table 1 Use of Survey Modes in German Market Research

	2004	2009	2014
PAPI	24%	13%	4%
CAPI	7%	6%	10%
CATI	44%	42%	37%
Mail	9%	7%	6%
CAWI	16%	32%	43%

Source: ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.[2] PAPI=Paper and Pencil Interview; CAPI=Computer Assisted Personal Interview; CATI=Computer Assisted Telephone Interview; Mail=Self-administered mailed questionnaires; CAWI=Computer Assisted Web Interviewing.

Table 2 Components of Survey Response Process and Cognitive Skills Required

Component	Specific Processes	Cognitive abilities
Comprehension	Attend to questions and instructions	Global Reading abilities
	Represent logical form of questions	
	Identify question focus (information sought)	
	Link key terms to relevant concepts	
Retrieval	Generate retrieval strategy and cues	Working Memory functioning:
	Retrieve specific, generic memories	- Spatial Updating Performance
	Fill in missing details	- Processing Speed
Judgment	Assess completeness and relevance of memories	Episodic Memory functioning
	Draw inferences based on accessibility	
	Integrate material retrieved	
	Make estimate based on partial retrieval	
Response	Map judgment onto response category	Fluid Intelligence
	Edit response	

Modified version of table in Tourangeau, Rips et al. (2000, p. 8).

[2] Data is provided on the following webpage of the ADM: https://www.adm-ev.de/zahlen/#c245 (date of retrieval: September 14, 2015)

Table 3 Assignment to MAIL and CAWI in 2012 by Survey Mode in 2009 (Household Level)

Participation in 2009 Survey	Assignment in 2012 Survey MAIL	CAWI	Total
PAPI Interview	17 (50%)	17 (50%)	34
CAPI Interview	41 (54.0%)	35 (46.0%)	76
MAIL Interview	230 (52.0%)	212 (48.0%)	442
CAWI Interview	194 (50.1%)	193 (49.9%)	387
Refusal of Interview	56 (46.7%)	64 (53.3%)	120
Not Yet in Sample	175 (51.6%)	164 (48.4%)	339
Total	713 (51.0%)	685 (49.0%)	1,398

Table 4 Survey Mode in 2012 by Assignment Groups (Household Level)

Assignment Group	Mode Choice in 2012 MAIL	CAWI	N
MAIL	611 (85.7%)	102 (14.3%)	713
CAWI	244(35.6%)	441 (64.4%)	685
N	855 (61.2%)	543 (38.8%)	1,398

Table 5 Rate of Compliance with Mode Assignment in 2012 by Survey Mode in 2009 (Household Level)

Choice in 2009 Survey	Assignment in 2012 Survey MAIL	CAWI
PAPI Interview	1.00	0.35
CAPI Interview	0.88	0.57
MAIL Interview	0.99	0.35
CAWI Interview	0.78	0.85
Refusal of Interview[1]	0.86	0.73
Not Yet in Sample	0.75	0.78
Total	611	441

Table 6 Respondents' Characteristics by Survey Mode in 2012

	CAWI	MAIL
Age (mean)	55.9	64.2
Male (%)	56%	45%
Secondary Education[a] (N, row %)	67%	55%
Mode 2009 (%[b])		
CAWI	38.1%	21.1%
Mail	14.4%	40.6%
Face-to-Face	5.7%	9.2%
Interview Duration (mean)	56.0	57.6
Conscientiousness (mean)	-0.05	0.03
Goal engagement (mean)	7.62	7.65
Goal disengagement (mean)	6.18	6.22
Working memory: Digit symbol test (mean)	0.21	-0.12
Working memory: Spatial updating (mean)	0.29	-0.17
Global reading score (mean)	0.19	-0.11
Fluid intelligence (mean)	0.35	-0.21
Episodic memory (mean)	0.15	-0.09

[a] "Secondary Education" refers to the German *Abitur*, comparable to the "General Certificate of Education" in England.
[b] The sum of percentages does not achieve the value of 100 since nonrespondents from 2009 and new recruits are not considered here.

Table 7a Means of Indicators of Data Quality by Survey Mode in 2012

	CAWI	MAIL	t-value from t-test
Social desirability	4.6	4.8	0.02***
Item nonresponse	0.03	0.03	4.91
Midpoint categories	0.19	0.21	2.97***
Extremeness	0.45	0.44	-1.78*
Person-reliability	0.08	0.10	0.90

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7b Means of Indicators of Data Quality by Cognitive Abilities

		Social desirability	Item Nonresponse	Midpoint categories	Extremeness	Person-Reliability
Speed processing: Digit symbol test	low	4.95	.03	.21	.44	.13
	high	4.51***	.03***	.19***	.44	.10
Working memory: Spatial updating	low	5.01	.03	.21	.44	.11
	high	4.57***	.02***	.20***	.44	.11
Global reading score	Low	4.90	.03	.20	.44	.10
	high	4.63***	.02***	.20*	.45	.13
Fluid intelligence	Low	5.03	.03	.21	.44	.10
	high	4.58***	.02***	.20***	.44	.12
Episodic memory	Low	4.89	.03	.21	.44	.10
	high	4.64***	.03***	.20***	.45	.12

Note: *p<0.1; **p<0.05; ***p<0.01. Significance test is based on t-values from t-tests.

Table 8a The Impact of Global Reading Abilities (GRA) on Response Quality

	Item nr	midpoint	extremeness	Person fit	Social desirability
CAWI	0.006***	-0.014**	0.017*	-0.023	-0.048
CAWI*GRA	0.002	0.007	-0.016	-0.026	-0.018
GRA	-0.002**	-0.007*	0.017***	0.033*	-0.048
Age	0.000*	0.000	0.000	0.004***	0.014***
Male	-0.009***	-0.009	0.001	0.017	-0.151*
High education	-0.006***	-0.002	-0.005	-0.003	0.029
F2f 2009	-0.001	0.006	0.006	-0.044	0.037
CAPI 2009	-0.002	-0.016	0.036	-0.098	0.176
CAWI 2009	-0.006**	0.006	0.007	-0.015	-0.233**
Refusal 2009	0.001	0.012	-0.011	-0.103*	-0.102
1. participation 2012	-0.001	-0.003	0.012	-0.061	-0.182
Answer time	-0.000**	0.000	-0.000	0.000	0.004**
Conscientiousness	-0.003***	-0.018***	0.038***	-0.025*	0.181***
Goal engagement	0.000	-0.049***	0.080***	0.068	0.526***
Goal disengagement	-0.000	-0.004	0.013	0.013	0.086
Time difference between measures	0.001	-0.012**	0.014	0.015	0.006
Constant	0.028***	0.331***	0.171***	-0.325**	2.221***
R^2	0.098	0.159	0.205	0.056	0.172
N	671	671	671	671	671

Table 8b_1 The Impact of Working Memory (Spatial Updating) on Response Quality

	Item nr	midpoint	extremeness	Person fit	Social desirability
CAWI	0.006***	-0.013**	0.016*	-0.012	-0.053
CAWI*Spatial U.	0.003	-0.006	0.013	-0.080***	-0.087
Spatial Updating	-0.004***	-0.000	-0.001	0.059***	0.045
Age	0.000	0.000	0.000	0.004***	0.014***
Male	-0.008***	-0.009	0.001	0.016	-0.152*
High education	-0.005***	-0.003	-0.001	-0.009	0.008
F2f 2009	-0.001	0.006	0.006	-0.033	0.047
CAPI 2009	-0.000	-0.017	0.040*	-0.122*	0.139
CAWI 2009	-0.005**	0.005	0.010	-0.025	-0.255**
Refusal 2009	0.001	0.011	-0.007	-0.102*	-0.109
1. participation 2012	-0.000	-0.004	0.013	-0.073*	-0.193
Answer time	-0.000**	0.000	-0.000	-0.000	0.004***
Conscientiousness	-0.003***	-0.018***	0.036***	-0.026*	0.188***
Goal engagement	-0.000	-0.051***	0.083***	0.069	0.517***
Goal disengagement	-0.000	-0.002	0.010	0.014	0.100
Time difference between measures	0.001	-0.012**	0.013	0.012	0.009
Constant	0.032***	0.336***	0.163***	-0.325**	2.192***
R^2	0.108	0.157	0.198	0.068	0.171
N	671	671	671	671	671

Table 8b_2 The Impact of Working Memory (Processing Speed) on Response Quality

	Item nr	midpoint	extremeness	Person fit	Social desirability
CAWI	0.006***	-0.012**	0.016*	-0.021	-0.052
CAWI*Processing Speed	0.003	0.001	0.009	-0.048*	0.013
Processing Speed	-0.004***	-0.013***	0.006	0.047**	-0.087
Age	0.000	-0.000	0.000	0.005***	0.012***
Male	-0.009***	-0.013**	0.005	0.023	-0.173*
High education	-0.006***	-0.002	-0.002	-0.001	0.029
F2f 2009	-0.000	0.008	0.006	-0.049	0.056
CAPI 2009	-0.002	-0.013	0.038*	-0.102	0.195
CAWI 2009	-0.006**	0.005	0.009	-0.013	-0.241***
Refusal 2009	0.001	0.012	-0.009	-0.101*	-0.096
1. participation 2012	-0.002	-0.004	0.014	-0.060	-0.182
Answer time	-0.000**	0.000	-0.000	0.000	0.004***
Conscientiousness	-0.003***	-0.018***	0.036***	-0.027*	0.186***
Goal engagement	0.000	-0.049***	0.081***	0.069	0.529***
Goal disengagement	0.000	-0.003	0.011	0.010	0.086
Time difference between measures	0.000	-0.013**	0.013	0.016	0.001
Constant	0.033***	0.360***	0.147***	-0.370**	2.378***
R^2	0.102	0.169	0.200	0.060	0.173
N	671	671	671	671	671

Table 8c The Impact of Episodic Memory Functioning (EMF) on Response Quality

	Item nr	midpoint	extremeness	Person fit	Social desirability
CAWI	0.006***	-0.012*	0.015	-0.019	-0.045
CAWI*EMF	0.004**	-0.006	0.017*	-0.031	-0.035
EMF	-0.004***	-0.004	-0.002	0.033*	-0.055
Age	0.000	0.000	0.000	0.004***	0.013***
Male	-0.009***	-0.011*	0.002	0.023	-0.168*
High education	-0.005***	-0.003	-0.001	-0.003	0.032
F2f 2009	-0.000	0.006	0.007	-0.045	0.048
CAPI 2009	-0.001	-0.015	0.040*	-0.107	0.207
CAWI 2009	-0.005**	0.006	0.011	-0.020	-0.231**
Refusal 2009	0.001	0.011	-0.008	-0.098*	-0.103
1. participation 2012	-0.001	-0.004	0.015	-0.065	-0.178
Answer time	-0.000**	0.000*	-0.000	-0.000	0.004***
Conscientiousness	-0.003***	-0.018***	0.037***	-0.028**	0.189***
Goal engagement	-0.000	-0.050***	0.083***	0.072	0.517***
Goal disengagement	-0.000	-0.002	0.009	0.010	0.092
Time difference between measures	0.001	-0.012**	0.013	0.014	0.005
Constant	0.032***	0.342***	0.163***	-0.345**	2.308***
R^2	0.110	0.161	0.200	0.056	0.173
N	671	671	671	671	671

Table 8d The Impact of Fluid Intelligence (FI) on Response Quality

	Item nr	midpoint	extremeness	Person fit	Social desirability
CAWI	0.007***	-0.010*	0.016*	-0.021	-0.020
CAWI*FI	0.002	-0.006	0.001	-0.055*	-0.145
FI	-0.005***	-0.006	0.008	0.064***	-0.005
Age	0.000	0.000	0.000	0.005***	0.013***
Male	-0.008***	-0.009	0.001	0.015	-0.142
High education	-0.005**	-0.001	-0.004	-0.011	0.038
F2f 2009	-0.000	0.006	0.007	-0.047	0.046
CAPI 2009	-0.000	-0.015	0.036	-0.125*	0.163
CAWI 2009	-0.004*	0.006	0.007	-0.035	-0.263**
Refusal 2009	0.002	0.012	-0.010	-0.018**	-0.116
1. participation 2012	0.001	-0.001	0.010	-0.084**	-0.177
Answer time	-0.000**	0.000	-0.000	0.000	0.004***
Conscientiousness	-0.003***	-0.018***	0.037***	-0.025*	0.184***
Goal engagement	-0.001	-0.052***	0.083***	0.073	0.496***
Goal disengagement	0.000	-0.002	0.010	0.012	0.101
Time difference between measures	0.001	-0.012**	0.013	0.015	0.014
Constant	0.033***	0.345***	0.162***	-0.345**	2.341***
R^2	0.121	0.163	0.198	0.069	0.174
N	671	671	671	671	671