Reliability and Completion Speed in Online Questionnaires under consideration of Personality

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

20 February 2017

DOI: 10.1016/j.paid.2017.02.015
Cite as:
Harms, C., Jackel, L., & Montag, C. (2017). Reliability and completion speed in online questionnaires under consideration of personality. Personality and Individual Differences, 111, 281–290. http://doi.org/10.1016/j.paid.2017.02.015

ABSTRACT

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

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

[*] *Corresponding author. Email address:* christopher.harms@uni-bonn.de
[a] *Department of Psychology, University of Bonn, Bonn, Germany*
[b] *Institute for Psychology and Education, Ulm University, Ulm, Germany*
[c] *Key Laboratory for NeuroInformation, Center for Information in Medicine, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China*

1 INTRODUCTION

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

While exploiting the benefits, less consideration goes into the question of data quality in practice. Past research has shown the equivalence of pen-and-paper and online application of personality questionnaire extensively in many different areas (Campos et al., 2011; Carlbring et al., 2007; Chuah et al., 2006). In contrast to traditional data collection, additional data on the participant and its response behavior is available. So-called "meta-data" on the user (e.g. used web-browser, operating system) and "para-data" (data on the process of filling in the questionnaire, e.g. reaction time, scrolling distance on the page, typing speed, etc.) can be very easily and objectively collected (Stieger and Reips, 2010) without interfering with the actual questionnaire. This allows researchers to include further behavioral and technical variables in their analysis, such as precisely measured response latencies (which would belong to the category of "para-data"). These meta- and para-data can be used as a first indicator of survey data quality (Furnham et al., 2013; Gummer and Rossmann, 2015; Heerwegh, 2003). In particular, when using any para-data to filter respondents in a sample, it is crucial for survey operators or researchers to understand the implications on both the quality of their data and on the validity of conclusions. Filtering respondents might otherwise lead to a bias in the sample and can reduce generalizability of the results. Thus, it seems a reasonable question to ask if para-data used to filter respondents such as response latencies are correlated with any of the constructs, which again are connected to the theory investigated. If, for example, response times in an online marketing survey were negatively correlated with household income, filtering respondents based on their individual response time and removing very fast participants would result in a sample with under-representation of households with low income. If household income was related to the research question in any direct or indirect way, conclusions based on the filtered data have to account for this bias. If not household income but personality dimensions were correlated with response time, the effect on the research question might be less direct but still relevant.

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

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

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

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

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

In general, it seems reasonable to assume an effect of Impulsiveness on the response behavior in questionnaires following the presented rationale: dysfunctional Impulsiveness "as a predisposition toward rapid, unplanned reactions to [...] external stimuli without regard to the negative consequences of these reactions [...]" (Moeller et al., 2001, as cited in Stanford et al., 2009, p. 385) should also play a role in self-report situations where questions are presented as external stimuli. As prior research has shown, impulsive subjects are, for example, faster in reaction-time experiments (Edman et al., 1983) and slower in reactions to a Stroop paradigm (Enticott et al., 2006). For speeded tests of cognitive ability Reeve (2007) highlighted the importance of functional Impulsiveness. With respect to response times for self-report questionnaires, in which participants do not have a time limit and cognitive ability is of minor relevance (at least compared to cognitive tests), mixed results have been found for a link between response latencies and Impulsiveness for paper-pencil questionnaires: Malle and Neubauer (1991) found no link between self-reported Impulsiveness and the response latencies in the Matching Familiar Figures Test (MFFT; Kagan et al., 1964), a test to assess cognitive tempo, in 65 undergraduate subjects. On the other hand, Moltó et al. (1993) used a 261-item self-report questionnaire comprising the I5 Impulsiveness Scale (Eysenck and Eysenck, 1978) and "other personality scales" (Moltó et al., 1993, p. 97) and tested the relationship between total response time and Impulsiveness. In a sample of 795 undergraduate students, faster participants had higher Impulsiveness scores in both male and female participants. Their decision to Median-split the sample based on response time, however, reduces their statistical power significantly (Cohen, 1983) and the dependent measure, total response time, was measured manually by noting the time questionnaires were handed in by the participants to the experimenter. Thus, the measurement was not precise and prone to different influences not related to the questionnaire completion. More recent research of online surveys did not directly address Impulsiveness or other common psychological personality constructs and their effect on response latencies or overall completion time of (online-)questionnaires (Furnham et al., 1998, 2013; Gummer and Rossmann, 2015; Yan and Tourangeau, 2008).

We seek to fill this gap in the literature with the present study and investigate how completion time in an online-questionnaire and personality dimensions are linked to each other. If survey operators or researchers are using cut-off values in completion time to filter the sample aiming for higher data quality, a correlation between completion time and personality would introduce a bias in the sample, that often goes unaccounted for in the data analysis. With the availability of server-side reaction times in our design, we aimed to investigate how Impulsiveness effects the completion time of a self-report personality questionnaire. In particular, we hypothesized that Impulsiveness would predict faster completion of the questionnaires, following the findings by Moltó et al. (1993).[1] As reaction times are possibly correlated with Age and Education as a proxy for cognitive ability (Yan and Tourangeau, 2008), we present data in- and excluding these variables as controls.

2 METHOD

2.1 Participants

A convenience sample of participants was recruited, mainly under-graduates at the University of Bonn and Ulm University. Participants studying psychology could receive course credit for their participation. 572 subjects started the questionnaire. As incomplete data sets were not included, N = 532 cases are used for subsequent analysis as they have finished the study and provided complete data sets. Sample descriptives are shown in Table 1.

Table 1: Sample descriptives

N	532
Age	M = 25.57 (SD = 6.30)
Female	436 (81.95%)
Completion time (seconds)	M = 751.43 (SD = 1318.17) Median = 554

2.2 Experimental Procedure

After reading and agreeing to the terms of the study, participants answered demographic questions, including Age, Gender and Educational level. Participants were informed that para-data were collected, but the nature of the data was not explicitly stated in order to not influence them in their natural behavior of completing the questionnaire. At no point of the study, partic-

1 As a reviewer noted, the opposite direction of the effect is also plausible: impulsive participants might more easily be distracted by cues outside the questionnaire and thus interrupt the completion more frequently. The used statistical approach does not imply any direction of the effect, i.e. both outcomes should be visible in the results.

ipants could provide their names. Therefore, participation was completely anonymous. The educational level was indicated by a choice with seven options representing school degrees (or no degree) in Germany.

The subsequent main self-report questionnaire consisted of 90 personality items (60 questions from NEO-FFI, 30 questions from Barratt's Impulsiveness Scale) in German language (Borkenau and Ostendorf, 1993; Preuss et al., 2008). Items were presented at random order across both questionnaires for each participant in a single-item-per-page design, where participants had to respond to the items on a 5-point Likert scale.[2] All data and para-data were automatically stored on the server, including time-stamps of starting and finishing the questionnaire. In the context of the present research, para-data comprises the overall completion time of the questionnaire, which is simply calculated by the difference between starting and finishing the questionnaire.

Students, who received course credits for their participation, entered a personal code at the end of the questionnaire that was stored independently from the questionnaire responses to maintain anonymity.

For the reliability analysis the sample was divided into five equally large groups based on participants' overall completion time. Although, we cannot be sure that participants worked diligently through questionnaires without taking breaks, we refrained from excluding any participants from the analysis, although some participants obviously took breaks (e.g. completion time of more than 25,000 s, i.e. more than 6.5 h). In the earlier study by Montag and Reuter (2008), a pre-experiment was carried out where it was reported that participants needed about 13 min to fill in 110 items (with a standard deviation of 2.76; p. 720). Such a pre-experiment was not conducted here and we did not pre-register any exclusion rule. But as detailed in the result section, the very slow or fast groups did not differ in terms of the internal consistencies. For the analysis of a relationship between Impulsiveness and completion time, we report results for a trimmed sample excluding the fastest and slowest 5%.[3]

2.3 Statistical Analyses

Internal consistency was measured using Cronbach's α (Cronbach, 1951) as a lower bound for the reliability and McDonald's ω (McDonald, 1999; Dunn et al., 2014) as an improved index of internal consistency. Further, measurement invariance was assessed through multi-group confirmatory

2 The BIS-11 items originally use 4-point Likert scales. As we were not looking for comparisons to existing norms, we chose to transform the response scale to a 5-point scale to use the same scales across both questionnaires. In our view this does not alter the general interpretation of the scores.

3 The decision is mainly arbitrary. Since we did not pre-register the study and using the collected data to find reasonable bounds confounds the results, readers should read these analyses with care.

factor analysis (Cheung and Rensvold, 2002; Hirschfeld and Von Brachel, 2014).

To analyze the effect of personality on completion time, we use a Bayesian approach instead of traditional Null-Hypothesis Significance Testing (NHST): We calculate a weighted likelihood ratio between our hypothesized model and a model containing only the covariates Age and Educational level, known as "Bayes factor" (BF). The calculation of the Bayes factor BF_{10} for linear regression models requires a prior distribution to be placed on the standardized coefficients β (Rouder and Morey, 2012). Different approaches for selecting a prior distribution on parameters exist. For the present study we chose a weakly informative prior known as Jeffreys-Zellner-Siow (JZS) prior (Liang et al., 2008; Rouder and Morey, 2012; Rouder et al., 2009). Rouder and Morey (2012) recommended the JZS prior as a "default prior" since it implements desirable theoretical properties in the Bayes factor (location and scale invariance, consistency and consistent in information, see p. 883). However, the JZS prior is subjectively in the sense that a scale parameter, s, can be chosen. To show the robustness of our analysis, a prior sensitivity analysis is detailed in Appendix B.

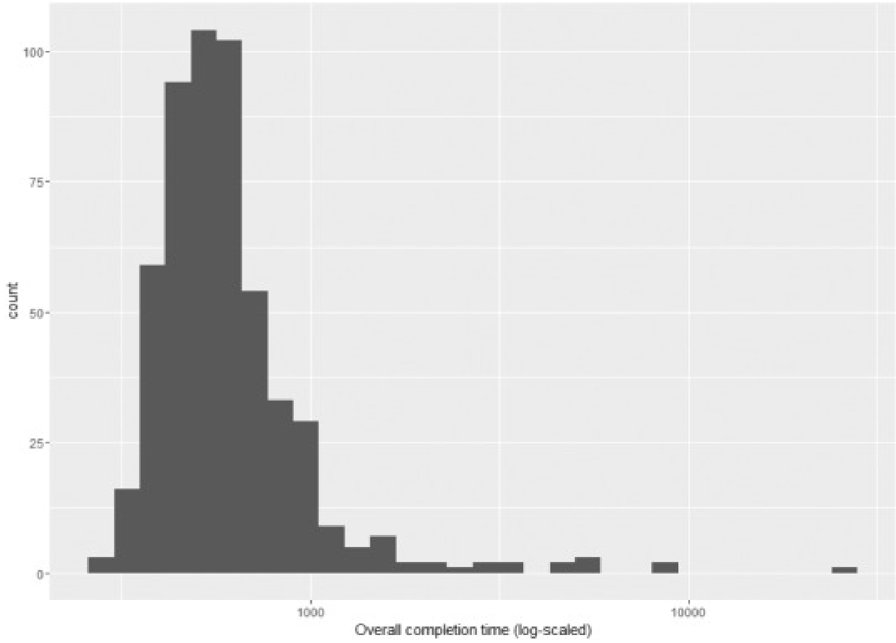
The Bayes factor, based on the JZS prior, can easily be computed using the `BayesFactor` package for R (Rouder et al., 2015). The Bayes factors are computed by the package using simulations, so an error estimate for the BF is given when reporting the results.

In the Bayesian framework, the Bayes factor is the degree to which a researcher should update his belief in either model after seeing the data (Dienes, 2016). It is a continuous quantification of the evidential value in the given data. While no arbitrary cut-offs are needed for the interpretation, readers might find the proposed bounds by Jeffreys (1961) useful: A $BF_{10} > 3$ is seen as "substantial evidence" in favor of the numerator model (Rouder et al., 2009; Wetzels and Wagenmakers, 2012), i.e. the data are at least three times more likely under the nominator model than under the denominator model.

It is a favorable property of the Bayes factor that it is also able to quantify the evidence against a hypothesized model, i.e. in favor of the denominator model. Thus, a $BF_{10} < 1/3$ can be interpreted as "substantial evidence" against the model in the numerator. Bayes factors between 1/3 and 3 hint at inconclusive evidence and researchers might want to increase sample size or re-evaluate their models.

Readers unfamiliar with the Bayesian approach can find a traditional hierarchical regression investigating the effects of personality on completion time in Appendix A. The following Results section reports only Bayes factors.

Figure 1: Histogram of completion time.



Note: Histogram for overall completion time of participants (log-scaled X-axis). Distribution is positively skewed ($\gamma_1 = 14.011$) as expected for reaction time data.

3 RESULTS

3.1 Age, Gender, Education and completion time

As expected, older participants needed longer to complete the survey ($r = 0.145$, $BF_{10} = 9.359$). The histogram in Fig. 1 shows that the distribution of completion time is positively skewed ($\gamma_1 = 14.011$) as to be expected with reaction time data (Anders et al., 2016; Gummer and Rossmann, 2015; Malhotra, 2008; Verdonck and Tuerlinckx, 2016; Wagenmakers, 2009; Wagenmakers and Brown, 2007; Yan and Tourangeau, 2008). A main effect of Education on completion time was not found (one-way Bayesian ANOVA, $BF_{10} = 0.117$; Rouder et al., 2012), evidence for or against an effect of Gender was inconclusive on the full sample (Bayesian t-Test for independent samples: $r = 0.707$, $BF_{10} = 1.276$; Rouder et al., 2009). In the following analyses, Education is included together with Age despite Education not showing evidence for an effect on the dependent variable in order to closely follow the model of Yan and Tourangeau (2008). Their study showed that both variables are potentially correlated to questionnaire completion time and thus are of potential interest to our research question. Further, Educational level is used as a rough proxy for cognitive ability, which is likely to influence the time a participant needs to complete the questionnaire.

Descriptive correlations between Age, personality scales and our dependent measure are presented in Table 2.

Table 2: Pearson correlations between speed and personality.

	CT	BIS-11	BIS-A	BIS-M	BIS-NP	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C
Age	0.145	0.031	-0.082	0.112	0.027	-0.117	-0.005	0.053	-0.153	0.068
Completion Time (CT)		-0.02	-0.029	0.009	-0.032	-0.005	-0.059	0.015	-0.056	0.013
BIS-11			0.716	0.846	0.848	0.176	0.02	0.062	-0.232	-0.617
Attentional Imp.				0.371	0.418	0.428	-0.094	0.009	-0.209	-0.473
Motor Imp.					0.62	-0.047	0.139	0.148	-0.238	-0.455
Non-Planning Imp.						0.087	-0.017	-0.02	-0.111	-0.568
NEO-N							-0.434	-0.014	-0.138	-0.202
NEO-E								0.135	0.215	0.191
NEO-O									0.084	-0.142
NEO-A										0.146

Note: N = 532.

3.2 Internal consistencies of NEO-FFI and BIS-11 depending on the different speed groups

Coefficient α and McDonald's ω for the whole sample, each speed group and all subscales are reported in Table 3 and Table 4 respectively. The speed groups have been constructed so that there are five equally sized groups. The resulting interval bounds for completion time are also noted in able 3 and Table 4. As can be seen from visual inspection, almost all scales have at least acceptable internal consistencies, $\alpha > 0.60$ and $\omega > 0.60$, with no decisive differences between groups based on the bootstrapped confidence intervals except for the "Non-Planning Impulsiveness" scale of the BIS-11 which shows lower internal consistencies in the fourth group.[4]

3.3 Confirmatory factor analysis for measurement invariance

While Coefficient α and McDonald's ω only assess internal consistency, Multi-Group Confirmatory Factor Analysis was conducted (Cheung and Rensvold, 2002; Hirschfeld and Von Brachel, 2014). Table 5 reports the fit statistics for the scales and the associated χ^2-tests for differences in model fit by adding model constraints.

Readers might note the fit indices for Model 1 (configural model) which are not very satisfying when considering Hu and Bentler (1999)'s proposed cutoff values for the Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA). Other researchers, however, have argued that these cutoff values are not suitable for the investigation of trait models (Beauducel and Wittmann, 2005; Marsh et al., 2004; Seib-Pfeifer et al., 2017) especially for complex personality traits – as is the case with Impulsiveness considered here.

For all but one scale, namely Conscientiousness, the tests for changes in model fit through the addition of constraints on the factor residuals are not significant (see row "Model 4" in Table 5). Except for the Conscientiousness scale ($\Delta\chi^2(48) = 84.224$, $p < .001$), the assumption of Measurement Invariance thus cannot be rejected.

3.4 Impulsiveness and completion time

Performing Bayesian model selection using default JZS priors on the standardized slope parameters β as proposed by Liang et al. (2008) and Rouder and Morey (2012), allows us to quantify the evidential value of our data in light of our hypothesized model. The model containing only covariates, Age and Education, yields $\mathrm{BF}_{10} = 7.138$ ($\pm 1.49\%$), when compared to an

4 Traditional significance tests for differences between the groups were not performed, since significance tests cannot yield helpful evidence in favour of the null hypothesis of no difference between the groups. Also, the number of tests would lead to a corrected alpha level that would nearly rule out any significant effects. We recommend the interpretation based on reported estimates and bootstrapped confidence intervals.

Table 3: Coefficient α for all included personality scales.

Group	N	BIS-11	Attentional	Motor	Non-Planning	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C
All	532	0.821	0.689	0.677	0.636	0.838	0.744	0.753	0.807	0.766
		[0.798; 0.842]	[0.647; 0.726]	[0.631; 0.718]	[0.585; 0.681]	[0.819; 0.856]	[0.712; 0.771]	[0.721; 0.781]	[0.779; 0.832]	[0.738; 0.792]
1 (fastest)	104	0.813	0.626	0.690	0.651	0.835	0.751	0.789	0.817	0.814
		[0.751; 0.860]	[0.514; 0.717]	[0.580; 0.782]	[0.544; 0.731]	[0.785; 0.875]	[0.667; 0.816]	[0.720; 0.845]	[0.760; 0.866]	[0.756; 0.859]
2	108	0.836	0.730	0.625	0.671	0.829	0.765	0.796	0.828	0.813
		[0.785; 0.883]	[0.631; 0.806]	[0.481; 0.734]	[0.565; 0.763]	[0.779; 0.868]	[0.691; 0.816]	[0.738; 0.842]	[0.767; 0.879]	[0.765; 0.851]
3	107	0.842	0.678	0.741	0.711	0.840	0.753	0.718	0.800	0.667
		[0.799; 0.876]	[0.588; 0.751]	[0.663; 0.803]	[0.619; 0.785]	[0.798; 0.874]	[0.693; 0.800]	[0.630; 0.792]	[0.734; 0.849]	[0.575; 0.743]
4	105	0.806	0.725	0.686	0.417	0.857	0.678	0.713	0.790	0.761
		[0.751; 0.849]	[0.648; 0.786]	[0.577; 0.765]	[0.241; 0.569]	[0.818; 0.888]	[0.574; 0.756]	[0.609; 0.786]	[0.709; 0.851]	[0.680; 0.818]
5 (slowest)	108	0.809	0.660	0.644	0.634	0.822	0.766	0.741	0.797	0.742
		[0.745; 0.858]	[0.544; 0.752]	[0.538; 0.732]	[0.517; 0.726]	[0.776; 0.860]	[0.701; 0.820]	[0.664; 0.799]	[0.730; 0.850]	[0.675; 0.800]

Note: Differences in group size are due to rounding. Intervals for overall completion time in seconds: Group 1 = [275; 430], Group 2 = [433; 509]; Group 3 = [511; 598]; Group 4 = [599; 757]; Group 5 = [759; 25645]. Intervals in brackets next to estimates are 95% Confidence Intervals for Coefficient Alpha based on 10,000 bootstrapping iterations.

Table 4: McDonald's ω for all included personality scales.

Group	N	BIS-11	Attentional	Motor	Non-Planning	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C
All	532	0.824 [0.801; 0.845]	0.719 [0.679; 0.753]	0.670 [0.618; 0.713]	0.640 [0.588; 0.685]	0.861 [0.843; 0.877]	0.771 [0.735; 0.799]	0.766 [0.733; 0.795]	0.812 [0.783; 0.836]	0.803 [0.776; 0.827]
1 (fastest)	104	0.816 [0.727; 0.869]	0.673 [0.567; 0.747]	0.657 [0.499; 0.764]	0.626 [0.255; 0.727]	0.862 [0.812; 0.901]	0.770 [0.678; 0.837]	0.794 [0.713; 0.851]	0.822 [0.756; 0.869]	0.839 [0.779; 0.882]
2	108	0.838 [0.779; 0.886]	0.751 [0.652; 0.822]	0.624 [0.422; 0.736]	0.679 [0.530; 0.771]	0.855 [0.805; 0.890]	0.788 [0.693; 0.840]	0.819 [0.763; 0.861]	0.831 [0.762; 0.881]	0.850 [0.800; 0.887]
3	107	0.842 [0.788; 0.877]	0.727 [0.643; 0.786]	0.745 [0.656; 0.805]	0.720 [0.625; 0.793]	0.867 [0.825; 0.897]	0.786 [0.707; 0.832]	0.733 [0.634; 0.801]	0.807 [0.737; 0.856]	0.711 [0.604; 0.781]
4	105	0.810 [0.726; 0.855]	0.758 [0.678; 0.816]	0.668 [0.491; 0.763]	0.475 [0.168; 0.606]	0.877 [0.836; 0.904]	0.720 [0.602; 0.803]	0.727 [0.606; 0.803]	0.799 [0.715; 0.857]	0.801 [0.727; 0.851]
5 (slowest)	108	0.815 [0.750; 0.866]	0.690 [0.576; 0.771]	0.664 [0.557; 0.743]	0.638 [0.490; 0.732]	0.843 [0.792; 0.879]	0.794 [0.711; 0.850]	0.754 [0.666; 0.816]	0.804 [0.731; 0.857]	0.783 [0.714; 0.834]

Note: Differences in group size are due to rounding. Intervals for overall completion time in seconds (same as in Table 3): Group 1 = [275; 430], Group 2 = [433; 509]; Group 3 = [511; 598]; Group 4 = [599; 757]; Group 5 = [759; 25645]. Intervals in brackets next to estimates are 95% Confidence Intervals for Omega based on 10,000 bootstrapping iterations.

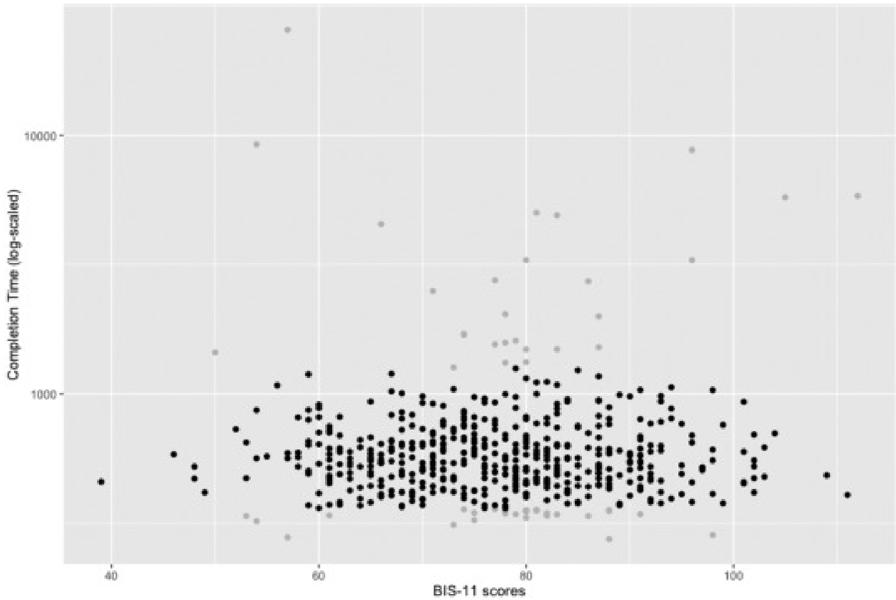
Table 5: Results of the Multi-Group Confirmatory Factor Analysis (MG-CFA).

		Attention	Motor	Non-Planning	NEO-N	NEO-E	NEO-O	NEO-A	NEO-C
Model 1	CFI	.857	.705	.712	.875	.694	.778	.809	.787
	RMSEA	.107	.108	.093	.102	.148	.099	.099	.125
Model 2	ΔRMSEA	.007	.008	< .001	.009	.006	.005	.009	.010
	$\Delta\chi^2$	41.761	42.505	78.198***	35.872	89.079***	57.907	29.934	40.236
	Δdf	28	40	40	44	44	44	44	44
Model 3	ΔRMSEA	.008	.002	.006	.006	.009	.007	.006	.006
	$\Delta\chi^2$	33.350	70.894	43.425	43.048	42.608	40.177	38.942	58.254
	Δdf	28	40	40	44	44	44	44	44
Model 4	ΔRMSEA	.010	.006	.007	.007	.009	.003	.004	.003
	$\Delta\chi^2$	27.615	48.496	36.273	39.638	36.695	62.411	51.921	84.224***
	Δdf	32	44	44	48	48	48	48	48
Model 5	ΔRMSEA	< .001	.001	.001	< .001	.001	.001	.001	< .001
	$\Delta\chi^2$	5.297	3.151	3.687	3.655	3.252	0.561	2.330	6.142
	Δdf	4	4	4	4	4	4	4	4

Note: Goodness-of-fit statistics for Multi-Group Confirmatory Factor Analysis to determine measurement invariance. Groups are defined by completion time as in Table 3 and Table 4. Model 1: Unconstrained model. Model 2: Constraints for equal loadings across all groups. Model 3: Constraints for equal loadings and intercepts. Model 4: Constraints for equal loadings, intercepts and residuals across all groups. Model 5: Constraints for equal loadings, intercepts, residuals and means. Null-hypothesis of measurement invariance is not rejected if χ^2-test for difference in model fit is not significant. Comparative Fit Index (CFI) and RMSEA are reported for Model 1 (configural model) and ΔRMSEA for the models adding constraints.

*** $p < .001$

Figure 2: Scatterplot of the data.



Note: Scatterplot showing distribution in the data. X-Axis are BIS-11 scores for participants and Y-axis shows participants' overall completion time (log-scaled). Grey dots represent participants removed for the analysis of the trimmed sample.

intercept-only model, indicating substantial evidence in favor of the model based on the work by Yan and Tourangeau (2008).

In fact, a post-hoc analysis using only Age as predictor for completion time yields $BF_{10} = 23.002$ ($\pm 0.01\%$), showing that the data contains the strongest evidence for an influence of only Age on the completion time.

Fig. 2 shows the distribution of the data on BIS-11 scores and completion time. To investigate the influence of Impulsiveness on the overall completion time, Bayesian model selection was again performed using default JZS priors. Comparing our hypothesized model including Age, Educational level and Impulsiveness with an alternative linear model only containing the covariates (Age and Education) yields $BF_{10} = 0.174$ ($\pm 1.77\%$), which indicates substantial evidence against our model (Jeffreys, 1961). In other words: our data is about 5.75 times more likely under the alternative model not including BIS-11.

Table 6 shows the posterior estimates of slope coefficients in both the covariate-only and the model including Impulsiveness.

3.4.1 *Exclusion of outliers*

If the fastest and slowest 5% are excluded from the analysis of Impulsiveness' effect on completion time, the sample size reduces to $N = 478$. The dependent variable is still positively skewed ($\gamma_1 = 1.176$). The trimmed sample does not contain evidence for the model including Age and Education when compared to a null-model ($BF_{10} = 0.063$, $\pm 0.89\%$). The evidence against our hypothesized model including Age, Education and Impulsive-

Table 6: Posterior Estimates for Regression Slopes

Variable	Covariate-only model	BIS & Covariates model
Mu	784.40	781.00
Age	35.34	34.68
No degree	3.88	16.29
Volks-/Hauptschulabschluss	-19.62	-60.10
Mittlere Reife	-95.59	-99.13
Fachabitur	195.00	214.40
Abitur	14.75	17.85
Fachhochschulabschluss	187.30	202.20
Hochschulabschluss	-285.80	-291.50
BIS-11		-3.26

Note: Unstandardized estimates for slope parameters. Estimated using posterior function from *BayesFactor* package with 1,000 iterations. Educational level was dummy-coded and each level is listed as variable in the table.

ness is still strong ($BF_{10} = 0.011$, $\pm 1.40\%$, when compared to an intercept-only null-model and $BF_{10} = 0.175$, $\pm 1.66\%$, when compared to the model containing the covariates). Therefore, despite the evidence against an effect of Age and Education in the trimmed sample, the model containing the BIS-11 score still performs worse.

Removing the covariates and testing the effect of only Impulsiveness on completion time yields also evidence against an effect ($BF_{10} = 0.102$, $\pm 0\%$).

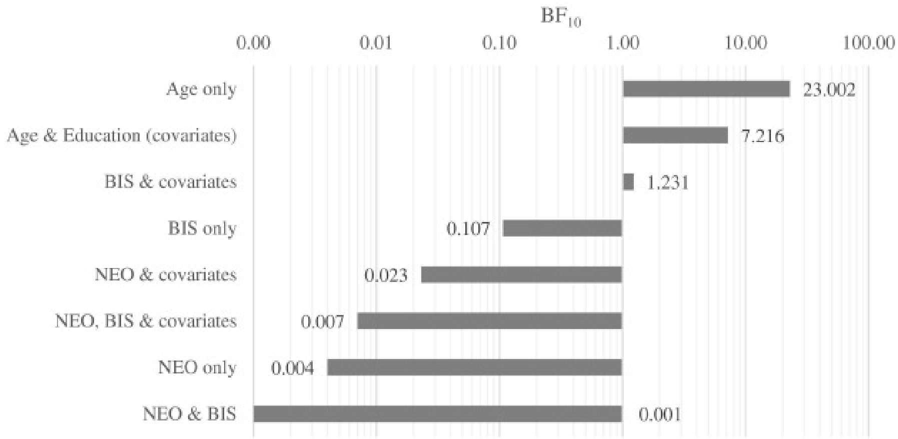
3.5 NEO-FFI and completion time

Although not primary focus of the present study, we also investigated if completion time was correlated with the NEO-FFI scales in the full sample. The results are thus of exploratory nature.

All correlations between completion time and personality are presented in Table 2. As there are notable correlations between the BIS-11 scale and Neuroticism ($r = 0.176$, $BF_{10} = 147.935$) and Conscientiousness ($r = -0.617$, $BF_{10} = 3.578 \times 10^{53}$), an empirical relationship between the Five-Factor Model of Personality and Impulsiveness can be assumed. Among others, this is also supported by the findings from Whiteside and Lynam (2001), who demonstrated that scales from BIS-11 and Impulsiveness-related facets from NEO-PI-R are loading on the same factors in a factor analysis. Dickman (1990) also showed that dysfunctional Impulsiveness is negatively correlated to Conscientiousness.

Using again Bayes factors to quantify the evidence in our data, none of the models yield stronger evidence than the model including only Age (see Fig. 3). In fact, there is substantial evidence against both a model including Age, Education, BIS-11 and NEO-FFI ($BF_{10} = 0.007$, $\pm 1.1\%$) and a model including only covariates and the NEO-FFI scales ($BF_{10} = 0.023$, $\pm 1.18\%$) when tested against an intercept-only model. Obviously, the evidence for

Figure 3: Model comparison.



Note: Values on x-Axis are Bayes factors (log-scaled axis; see Methods section for details), larger values mean more evidence for model in data. $BF_{10} = 1$: data are equally likely under both models. Covariates are Age and dummy-coded Educational level. Denominator is Intercept-only model as not all models included the covariates (age and dummy-coded educational level).

these two models is even weaker when controlling for covariates, i.e. testing against the model including only Age and Education as predictors ($BF_{10} = 0.001$, $\pm 1.37\%$, and $BF_{10} = 0.003$, $\pm 1.43\%$, respectively).

4 GENERAL DISCUSSION

We were able to replicate the findings by Montag and Reuter (2008) for apparent similarities in internal consistencies for different response speeds in our more precise design. Further, a confirmatory factor analysis shows that measurement invariance can be assumed for most scales with Conscientiousness being the exception. Further research is required to investigate the robustness of these findings, especially regarding the poor fit indices of the configural models. In general, we conclude that the time participants need to complete a self-report questionnaire by itself does not affect reliability measures. This replication should assure researchers and survey operators that response time by itself is not an indicator of data quality (at least in terms of internal consistencies). While control for content-sensitive measures of data quality still is important (e.g. length of text in open ended questions or suspicious patterns in Likert-scale questions), a decision based solely on respondents' completion time does not seem adequate in light of our findings.

While we were able to replicate the effects of Age on survey completion time at least in the full sample, the present study indicated evidence against our model including Impulsiveness. This leads us to the conclusion that dysfunctional Impulsiveness is not a relevant personality construct for predicting the time a participant needs to complete a self-report questionnaire. Further, our data also shows evidence against an influence of any of the Five Factor personality scales on completion speed. Confirmatory analyses in future studies are needed to back up our results.

One reviewer noted that Functional Impulsiveness (Dickman, 1990) instead might have an influence on the response time as participants may implicitly feel they should rush through the questionnaire. While we were not able to test this implied hypothesis, future studies might include information on participants' motivation and thinking while filling out the questionnaire.

Our study is limited by two factors:

First, our sample consists mainly of students who are familiar with questionnaires in general and personality self-report questions in particular. This also leads to little variation in the Education variable, which might explain why we did not find any effect of Education on completion time in contrast to prior findings. While online pools for participants are available, the question of data-quality of e.g. MTurk samples is still under discussion (e.g. Rouse, 2015). Future studies should, however, aim at more representative samples. Second, some participants clearly engaged in activities outside answering the questionnaire leading to completion times well above 1 h (median completion time around 9 min). Since we did not specify any exclusion rules before the data analysis, we decided to include the whole sample in our primary analysis. As the analysis of a trimmed sample excluding the fastest and slowest 5% revealed, the effect of Age is not robust in our study. Despite this, the evidence in the data is still strong against an effect of Impulsiveness – in either direction.

A final thing to be considered is how completion speed relates to internal consistencies, when participants of a study are "baited" with a monetary incentive. Here, participants might be more willing to just click through the questionnaires. Therefore, future studies should be conducted under considerations of different incentives compared to none (for an overview on incentives in web-surveys see Bosnjak and Tuten (2003) or Göritz (2006)). Note, however, that the study by Montag and Reuter (2008) used a lottery incentive and also did not find a link between completion time and internal consistencies.

To better understand participants' behavior in online studies, client-side technologies, such as JavaScript, can be used to measure further para-data. Especially, how often and how long a participant moves to a different browser window or is inactive on the questionnaire, should yield further insights. While our analysis does not support this hypothesis and our research design did not allow for further investigation of this question, researchers might be interested in the effect of Impulsiveness on the number of pauses in the survey. Some research already has been done in this regard without taking personality into consideration (Stieger and Reips, 2010). A comprehensive model of response times in web surveys, however, does not exist. Models from cognitive psychology such as the Diffusion Decision Model (Ratcliff, 1993) could be used as a basis for a model that can be applied to settings outside laboratories. Any such model has to include a variety of different influences, including psychological aspects (e.g. cognitive processes, mental ability) as well as characteristics of the questionnaire itself (e.g. effort needed to answer the questions, type of answers scales).

Appendix

A RESULTS OF CLASSICAL HIERARCHICAL REGRESSION

In addition to the Bayesian analyses performed in this study, we report our findings using classical methods using Null-Hypothesis Significance Testing (NHST) as most readers will be more familiar with this approach. As there is an ongoing debate on the use and misuse of p-values and Bayes factors alike, we report the traditional results as well. As detailed in the Method section, we believe that in the context of our study, the Bayesian approach is better suited to quantify evidence in favor of the "null". Therefore, we chose to include the Bayes factor analyses in the central sections of the paper and add the traditional analyses as appendix in order to make the article more readable.

Age and completion time are significantly correlated ($r = 0.145$, $p < 0.001$, 95% CI $[0.061; 0.227]$) in the full sample ($N = 532$), but not in the trimmed sample ($r = 0.012$, $p = 0.792$, 95% CI $[-0.078; 0.102]$, $N = 478$). The main effect of Education on completion time was not significant (one-way ANOVA, $F(6, 525) = 1.038$, $p = 0.400$). No significant difference between male and female participants with respect to completion time was observed (Welch's two sample t-test, $t(98.166) = -1.189$, $p = 0.237$).

The evidence in favor of a significant correlation between BIS-11 (dysfunctional) Impulsiveness and both Neuroticism ($r = 0.176$, $p < 0.0001$) and Conscientiousness ($r = -0.617$, $p < 0.0001$) is supported also by traditional significance testing.

The results of the analyses for Impulsiveness and the Five-Factor Model of Personality were combined in a hierarchical regression. In the first step, Age and Education were entered as predictors ($R^2 = 0.026$, $p = 0.001$) showing the significant effect of Age on completion time ($\beta = 0.165$, $p < 0.001$). Second, the NEO-FFI scales were entered ($R^2 = 0.031$, $\Delta R^2 = 0.005$, $p = 0.781$), showing no significant effect for any of the five scales. As the third step the global BIS-11 scale was added ($R^2 = 0.031$, $\Delta R^2 = 0.001$, $p = 0.564$) also showing no significant effects for any scale in the prediction of completion time. See Table 7 for detailed hierarchical regression results.

Table 7: Results of hierarchical regression.

Variable	Block 1 β	Block 2 β	Block 3 β
Age	0.165***	0.155**	0.158**
Education	-0.075	-0.074	-0.075
NEO-N		-0.017	-0.012
NEO-E		-0.064	-0.055
NEO-O		0.027	0.025
NEO-A		-0.027	-0.032
NEO-C		0.024	0.003
BIS-11			-0.034
R^2	0.026**	0.031	0.031
ΔR^2		0.005	0.001

Note: $N = 532$. ** $p = 0.001$. *** $p < 0.001$

B PRIOR SENSITIVITY ANALYSIS

Bayes factors use prior distributions to weight the model likelihoods. The proposed default JZS priors, which were used in this article, provide some useful properties to the resulting Bayes factors (Rouder and Morey, 2012). However, it still allows researchers to tune the priors to include information on the effect under investigation using a scale parameter. In the context of the analyses presented here, a researcher might pose different expectations on the size of the effects (i.e. the standardized slope parameters in the regression). All analyses reported in the result section use the default value of $s = \frac{1}{4}\sqrt{2}$. If we had reason to expect smaller effect sizes, a smaller s could be chosen and *vice versa*.

To underline the robustness of our analysis, we performed a sensitivity analysis. We calculated the Bayes factor for varying values of s comparing our hypothesized model including Impulsiveness to the model including only Age and Education. The result is shown in Fig. 4. For any value of s the resulting Bayes factor is smaller than 1 and even smaller than 1/3: Our conclusions thus seem to be robust against any reasonable scale of the JZS prior distribution.

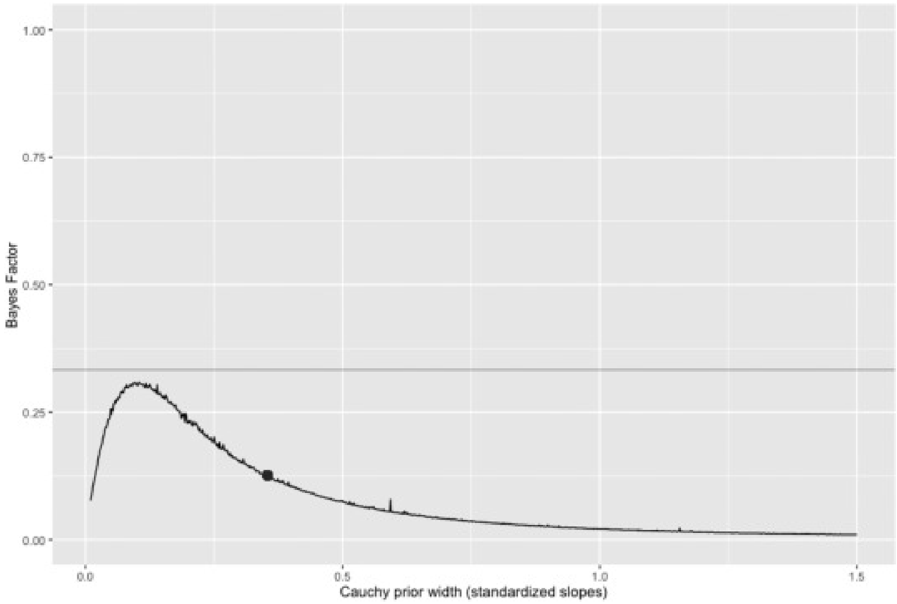
CONFLICT OF INTEREST

None.

CONTRIBUTIONS

CH, LJ and CM planned and designed the study. CH, LJ and CM carried out the data collection. CH carried out the statistical analysis and drafted

Figure 4: Robustness Check of JZS prior scale.



Note: Sensitivity analysis of Bayes Factor BF_{10} for comparing the hypothesized model with the model including Age and Education. The X-axis represents different values for the prior scale parameter. The black dot denotes the default Bayes Factor for $s = \frac{1}{4}\sqrt{2}$. Our interpretation of the Bayes Factors is considered robust as BF_{10} is always smaller 1/3 (grey line), showing consistently evidence against our hypothesized model including the Impulsiveness scale.

the manuscript. CM gave additional input with respect to the analysis and interpretation of the data. Finally, he worked over the initial draft written by CH.

ACKNOWLEDGEMENT

The position of Christian Montag is funded by a Heisenberg grant awarded to him by the German Research Foundation (DFG, MO2363/3-1).

REFERENCES

Anders, R., Alario, F.-X., and Van Maanen, L. (2016). The Shifted Wald Distribution for Response Time Data Analysis. *Psychological Methods*, 21(3):309–327.

Bassili, J. N. (1993). Response Latency Versus Certainty as Indexes of the Strength of Voting Intentions in a Cati Survey. *Public Opinion Quarterly*, 57(1):54.

Bassili, J. N. and Fletcher, J. F. (1991). Response-Time Measurement in Survey Research: A Method for CATI and a New Look at Nonattitudes. *Public Opinion Quarterly*, 55(3):331.

Beauducel, A. and Wittmann, W. W. (2005). Simulation Study on Fit Indexes in CFA Based on Data With Slightly Distorted Simple Structure. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(1):41–75.

Borkenau, P. and Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar:(NEO-FFI); nach Costa und McCrae*. Hogrefe, Göttingen.

Bosnjak, M. and Tuten, T. L. (2003). Prepaid and Promised Incentives in Web Surveys: An Experiment. *Social Science Computer Review*, 21(2):208–217.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1):3–5.

Campos, J. A. D. B., Zucoloto, M. L., Bonafé, F. S. S., Jordani, P. C., and Maroco, J. (2011). Reliability and validity of self-reported burnout in college students: A cross randomized comparison of paper-and-pencil vs. online administration. *Computers in Human Behavior*, 27(5):1875–1883.

Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L.-G., and Andersson, G. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, 23(3):1421–1434.

Cheung, G. W. and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2):233–255.

Chuah, S. C., Drasgow, F., and Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality*, 40(4):359–376.

Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, 7(3):249–253.

Costa, P. T. J. and McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Psychological Assessment Resources, Odessa, FL.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.

Davis, K. L., Panksepp, J., and Normansell, L. (2003). The Affective Neuroscience Personality Scales: Normative Data and Implications. *Neuropsychoanalysis*, 5(1):57–69.

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology*, 58(1):95–102.

Dienes, Z. (2016). How Bayes factors change scientific pratice. *Journal of Mathematical Psychology*, pages 1–29.

Dunn, T. J., Baguley, T., and Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3):399–412.

Edman, G., Schalling, D., and Levander, S. (1983). Impulsivity and speed and errors in a reaction time task: A contribution to the construct validity of the concept of impulsivity. *Acta Psychologica*, 53(1):1–8.

Egan, V., Deary, I., and Austin, E. (2000). The NEO-FFI: emerging British norms and an item-level analysis suggest N, A and C are more reliable than O and E. *Personality and Individual Differences*, 29(5):907–920.

Enticott, P. G., Ogloff, J. R., and Bradshaw, J. L. (2006). Associations between laboratory measures of executive inhibitory control and self-reported impulsivity. *Personality and Individual Differences*, 41(2):285–294.

Eysenck, S. B. G. and Eysenck, H. J. (1978). Impulsiveness and Venturesomeness: Their Position in a Dimensional System of Personality Description. *Psychological Reports*, 43(3f):1247–1255.

Furnham, A., Forde, L., and Cotter, T. (1998). Personality scores and test taking style. *Personality and Individual Differences*, 24(1):19–23.

Furnham, A., Hyde, G., and Trickey, G. (2013). On-line questionnaire completion time and personality test scores. *Personality and Individual Differences*, 54:716–720.

Göritz, A. S. (2006). Incentives in Web studies: Methodological issues and a review. *International Journal of Internet Science*, 1(1):58–70.

Gummer, T. and Rossmann, J. (2015). Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review*, 33(2):217–234.

Heerwegh, D. (2003). Explaining Response Latencies and Changing Answers Using Client-Side Paradata from a Web Survey. *Social Science Computer Review*, 21(3):360–373.

Hirschfeld, G. and Von Brachel, R. (2014). Multiple-Group confirmatory factor analysis in R - A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation*, 19(7):1–11.

Holden, R. R. (1995). Response latency detection of fakers on personnel tests. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 27(3):343–355.

Holden, R. R. (1998). Detecting Fakers on a Personnel Test: Response Latencies Versus a Standard Validity Scale. *Journal of Social Behavior and Personality*, 13(2):387.

Holden, R. R. and Hibbs, N. (1995). Incremental Validity of Response Latencies for Detecting Fakers on a Personality Test. *Journal of Research in Personality*, 29(3):362–372.

Hu, L. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1):1–55.

Jeffreys, H. (1961). *Theory of Probability*. Clarendon Press, Oxford, 3. ed. edition.

Kagan, J., Rosman, B. L., Day, D., Albert, J., and Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs: General and Applied*, 78(1):1–37.

Komar, S., Komar, J. A., Robie, C., and Taggar, S. (2010). Speeding Personality Measures to Reduce Faking. *Journal of Personnel Psychology*, 9(3):126–137.

Körner, A., Drapeau, M., Albani, C., Geyer, M., Schmutzer, G., and Brähler, E. (2008). Deutsche Normierung des NEO-Fünf-Faktoren-Inventars (NEO-FFI) German Norms for the NEO-Five Factor Inventory. *Zeitschrift für Medizinische Psychologie*, 17(2):133–144.

Liang, F., Paulo, R., Molina, G., Clyde, M. a., and Berger, J. O. (2008). Mixtures of g Priors for Bayesian Variable Selection. 103(481):410–423.

Malhotra, N. (2008). Completion Time and Response Order Effects in Web Surveys. *Public Opinion Quarterly*, 72(5):914–934.

Malle, B. F. and Neubauer, A. C. (1991). Impulsivity, reflection, and questionnaire response latencies: No evidence for a broad impulsivity trait. *Personality and Individual Differences*, 12(8):865–871.

Malloy-Diniz, L., Fuentes, D., Borges Leite, W., Correa, H., and Bechara, A. (2007). Impulsive behavior in adults with attention deficit/ hyperactivity disorder: Characterization of attentional, motor and cognitive impulsiveness. *Journal of the International Neuropsychological Society*, 13(04):693–698.

Marsh, H. W., Hau, K.-T., and Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3):320–341.

Maschke, P. (1989). Die Bearbeitungszeit von Persönlichkeitsfragebögen in der Eignungsauswahl: Ein Indikator für Verfälschung? *Zeitschrift für Differentielle und Diagnostische Psychologie*, 10(2):121–127.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Miller, E., Joseph, S., and Tudway, J. (2004). Assessing the component structure of four self-report measures of impulsivity. *Personality and Individual Differences*, 37(2):349–358.

Moltó, J., Segarra, P., and Avila, C. (1993). Impulsivity and total response speed to a personality questionnaire. *Personality and Individual Differences*, 15(1):97–98.

Montag, C. and Reuter, M. (2008). Does Speed in Completing an Online Questionnaire Have an Influence on Its Reliability? *Cyberpsychology & Behavior*, 11(6):719–721.

Patton, J. H., Stanford, M. S., and Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, 51(6):768–774.

Preuss, U. W., Rujescu, D., Giegling, I., Watzke, S., Koller, G., Zetzsche, T., Meisenzahl, E. M., Soyka, M., and Möller, H. J. (2008). Psychometrische Evaluation der deutschsprachigen Version der Barratt-Impulsiveness-Skala. *Der Nervenarzt*, 79(3):305–319.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299:172–179.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3):510–532.

Reeve, C. L. (2007). Functional Impulsivity and Speeded Ability Test Performance. *International Journal of Selection and Assessment*, 15(1):56–62.

Rouder, J. N., Love, J., Marwick, B., and Morey, R. (2015). BayesFactor: 0.9.12-2 CRAN.

Rouder, J. N. and Morey, R. D. (2012). Default Bayes Factors for Model Selection in Regression. *Multivariate Behavioral Research*, 47(6):877–903.

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5):356–374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2):225–237.

Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior*, 43:304–307.

Seib-Pfeifer, L.-E., Pugnaghi, G., Beauducel, A., and Leue, A. (2017). On the replication of factor structures of the Positive and Negative Affect Schedule (PANAS). *Personality and Individual Differences*, 107:201–207.

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., and Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, 47(5):385–395.

Stieger, S. and Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, 26(6):1488–1495.

Verdonck, S. and Tuerlinckx, F. (2016). Factoring out nondecision time in choice reaction time data: Theory and implications. *Psychological Review*, 123(2):208–218.

Voas, R. B. (1957). Personality Correlates of Reading Speed and the Time Required to Complete Questionnaires QUESTIONNAIRES. *Psychological Reports*, 3(3):177–182.

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, 21(5):641–671.

Wagenmakers, E.-J. and Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114(3):830–841.

Wetzels, R. and Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19(6):1057–1064.

Whiteside, S. P. and Lynam, D. R. (2001). The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, 30(4):669–689.

Yan, T. and Tourangeau, R. (2008). Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22(1):51–68.