

# Item-pair measures of acquiescence: the artificial inflation of socially desirable responding

Vaka Vésteinsdóttir<sup>a,b,c</sup>, Ragnhildur Lilja Asgeirsdóttir<sup>a,b</sup>, Ulf-Dietrich Reips<sup>c</sup>  
and Fanney Thorsdóttir<sup>a,b</sup>

<sup>a</sup>Quantitative Psychology, Department of Psychology, University of Iceland, Reykjavik, Iceland; <sup>b</sup>RAHÍ: Rannsóknarsetur í aðferðafræði við Sálfræðideild Háskóla Íslands/Methodology Research Center, University of Iceland, Reykjavik, Iceland; <sup>c</sup>Psychological Methods, Assessment, & iScience, Department of Psychology, University of Konstanz, Konstanz, Germany

## ABSTRACT

The purpose of this study was to evaluate the role of socially desirable responding in an item-pair measure of acquiescence from the Big Five Inventory. If both items in an item-pair have desirable content, the likelihood of agreeing with both items is increased, and consequently, the type of responding that would be taken to indicate acquiescence. In Study I, item content desirability was evaluated for each of the 32 items belonging to the item-pairs in two samples of 214 and 68 university students. The item-pair desirability was then correlated with the percentage of respondents who agreed with both items in a separate sample of 895 students. Results showed a substantial correlation between item-pairs' desirability and the percentage of estimated acquiescence, indicating an inflation of acquiescence when item-pairs have desirable content. The finding was further supported by Study II, in which acquiescence and item difficulty, assessed with cognitive interviews, were unrelated.

## KEYWORDS

Socially desirable responding; acquiescence; Big Five Inventory (BFI); response bias

## 1. Introduction

When assessing personality with self-report measures such as personality questionnaires the underlying assumption is that people's responses are based on their self-evaluation of how well the content of the items describes them as a person. Respondents are assumed to take several steps before deriving a response to each question. These steps involve understanding the question, retrieving the relevant information, making a judgment, and finally mapping the judgment on to the response scale (see Tourangeau et al., 2000 for a further outline of the response process).

It is however well-known that respondents do not always go through these steps before giving a response (Tourangeau et al., 2000) and sometimes base their answers on something other than the specific item content, resulting in biased responses (Paulhus, 1991). Response biases have been shown to affect measures of personality (see e.g. Ellingson et al., 1999; Holden, 2007; Soto et al., 2008; Viswesvaran & Ones, 1999). The two most well-known response biases are acquiescence and socially desirable responding (e.g. Paulhus, 1991; Weijters & Baumgartner, 2012). Acquiescence is defined as the tendency of respondents to agree with items regardless of item content (Weijters & Baumgartner, 2012) and is often thought to be the consequence of insufficient processing of the item content. Gilbert (1991) proposed that when people are presented with an idea, their initial reaction is to accept the idea and take it to be true, that is, that the first step in processing an idea, the

**CONTACT** Vaka Vésteinsdóttir [vakave@hi.is](mailto:vakave@hi.is) Quantitative Psychology, Department of Psychology, University of Iceland, Nýi Garður, Office 114, Sæmundargata 12, 102 Reykjavík, Iceland

comprehension stage, involves acceptance. Acceptance is thus always the first reaction, and rejection of an idea is always secondary and requires more cognitive effort. Therefore, when people are distracted, tired or uncertain about the content of a message they may fail to go beyond the initial response of acceptance. This idea of acquiescence being a product of limited processing is in line with how many others have approached this response bias, as acquiescence is thought to increase with uncertainty regarding item content, item difficulty and cognitive demand (see e.g. Cronbach, 1946; Knowles & Nathan, 1997; Krosnick, 1991).

Other ideas about what causes acquiescence have been put forth and focus mostly on the respondent's reaction to the social situation (see Krosnick, 1991). Acquiescence has, for example, been seen as a manifestation of a personality type where the so-called *yeasayers* are those who are more responsive to social situations (Couch & Keniston, 1960). Such explanations have however been refuted on the grounds of the absence of a correlation between measures of social desirability and acquiescence. For example, Knowles and Nathan (1997) found no correlation between a measure of acquiescence and the impression management scale from the Balanced Inventory of Desirable Responding (BIDR) and concluded that acquiescence was not caused by social concerns of the respondent. However, correlating social desirability scales with measures of acquiescence assumes that those who have a stronger tendency to respond in a socially desirable manner will be more prone to agreeing with items in general, regardless of the item content, and thus overlooks the connection between people's desire to present themselves favorably and the desirability of the item content. That is, if an item has undesirable content it is unlikely that the perceived socially desirable response will be agreeing with the item content instead of denying it. Therefore, although the absence of correlation between measures of acquiescence and social desirability scales can be taken as evidence that acquiescent responding is not due to the person's desire to give a good impression by generally agreeing with items, it does not rule out the possibility of socially desirable responding affecting measures of acquiescence.

This relates to how both acquiescence and social desirability are measured. Acquiescence is difficult to evaluate unless a questionnaire contains reversed items, as acquiescence cannot be separated from meaningful agreement with item content when all items are keyed in the same direction. Therefore, acquiescence is best assessed through item-pairs containing items with opposite implications, where agreement with both items would indicate acquiescent responding (see e.g. Knowles & Nathan, 1997; Weijters & Baumgartner, 2012). The question of whether socially desirable responding influences agreement with both items in an item-pair is, as explained above, not a question of the respondents' general desire to present themselves favorably, but a question of the desirability of the content of each item in the item-pair. For example, if both items in an item-pair have desirable content, despite having opposite implications, then a socially desirable response would be taken to indicate acquiescence.

It is therefore important to evaluate the desirability of the item content instead of the respondents' tendency to give desirable responses. This approach to measuring social desirability derives from Edwards (1957) who measured what he called Social Desirability Scale Value (SDSV), by asking respondents to rate the desirability of an item's content. Using this measure of social desirability, Edwards showed that the rated SDSV of items was highly correlated with endorsement frequency of items in a different sample from the sample that produced the SDSV ratings of the items. Therefore, given the connection between SDSV and endorsement frequency, the possible influence of socially desirable responding upon measured acquiescence can be evaluated by estimating the SDSV of each item in an item-pair, to test whether agreement between the two items is likely to be influenced by socially desirable responding.

The distinction between socially desirable responding and acquiescence is an important one as the two response biases are thought to be caused by different cognitive processes. Socially desirable responding is presumed to be caused by editing of responses during the response selection stage of the answering process, just before reporting (Tourangeau & Rasinski, 1988; Tourangeau et al., 2000). Respondents are thus assumed to have carefully evaluated the content of the item before

reporting in a socially desirable manner whereas acquiescent responding is presumed to be caused by lack of cognitive processing as acquiescence has been shown to increase with uncertainty regarding item content, item difficulty, and cognitive demands (see e.g. Cronbach, 1946; Knowles & Nathan, 1997; Krosnick, 1991).

The purpose of this research was to test whether acquiescence measured with item-pairs from a personality measure was affected by socially desirable responding. For this purpose, we used the item-pairs from the Big Five Inventory (BFI), identified by Soto et al. (2008) as being comprised of items with opposite implications. Based on the correspondence between item content desirability and item endorsement, we hypothesized in Study I that the higher the averaged SDSV of both items in a pair the more often participants would agree with both items. In Study II we assessed the difficulty of the items to provide further support for the findings of Study I.

## 2. Study I: evaluation of socially desirable responding in a measure of acquiescence

### 2.1. Method

#### 2.1.1. Participants and procedure

**2.1.1.1. The Big Five Inventory (BFI).** A link to an internet survey containing the BFI in Icelandic was sent out by the Student Registry at the University of Iceland to 8,141 undergraduate university students, of which 849 respondents answered at least one of the BFI item-pairs (i.e. both items in at least one pair). Of the 849 respondents whose responses were included in the analysis, 800 participants reported their gender: 608 females, 189 males and 3 marking the category 'Other'. Participants' age ranged from 18 to 67, with the mean age of 26 years ( $SD = 8.9$ ) for the 774 who reported their age.

**2.1.1.2. Social Desirability Scale Value (SDSV).** After the link to the first survey had been deactivated, the Student Registry at the University of Iceland sent out a second link to 7,209 undergraduate university students. The link contained a survey asking respondents to rate the desirability of the BFI items (and one other survey, not relevant to the current study, with random assignment to each survey). A total of 185 participants rated both items in at least one item-pair and were thus included in the analysis (this sample will hereafter be referred to as the UI sample). Of the 170 participants who reported their gender, 37 were male and 133 were female. The age of participants ranged from 18 to 67 years, with the mean age of 27 years ( $SD = 10.1$ ) for the 166 participants who reported their age.

Due to the possible overlap between respondents from the University of Iceland who answered the BFI (first survey) and who rated the SDSV of the BFI items (second survey), a second rating of SDSV was obtained from undergraduate students at Reykjavík University. The survey was sent out to 258 undergraduate students from either business or psychology (this sample will hereafter be referred to as the RU sample). Sixty-eight respondents gave an answer to at least one of the BFI item-pairs, of which 14 were male, 53 female and one did not report gender. Participants age ranged from 19 to 45, with the average age of 23 years ( $SD = 4.7$ ).

#### 2.1.2. Measures

**2.1.2.1. The Big Five Inventory (BFI).** The Big Five Inventory (BFI) is a five-factor measure of personality, consisting of 44 items, measuring Extraversion (8 items), Agreeableness (9 items), Conscientiousness (9 items), Neuroticism (8 items), and Openness (10 items) (John et al., 1991). The BFI was presented on a five-point response scale, with the following response options, coded from one to five: 'Disagree strongly', 'Disagree', 'Neither agree nor disagree', 'Agree' and 'Agree strongly'.

**2.1.2.2. Acquiescence.** Previous research (Soto et al., 2008) has identified 16 BFI item-pairs with 'opposite implications for personality' (pp. 723), based on item content and interitem correlation.

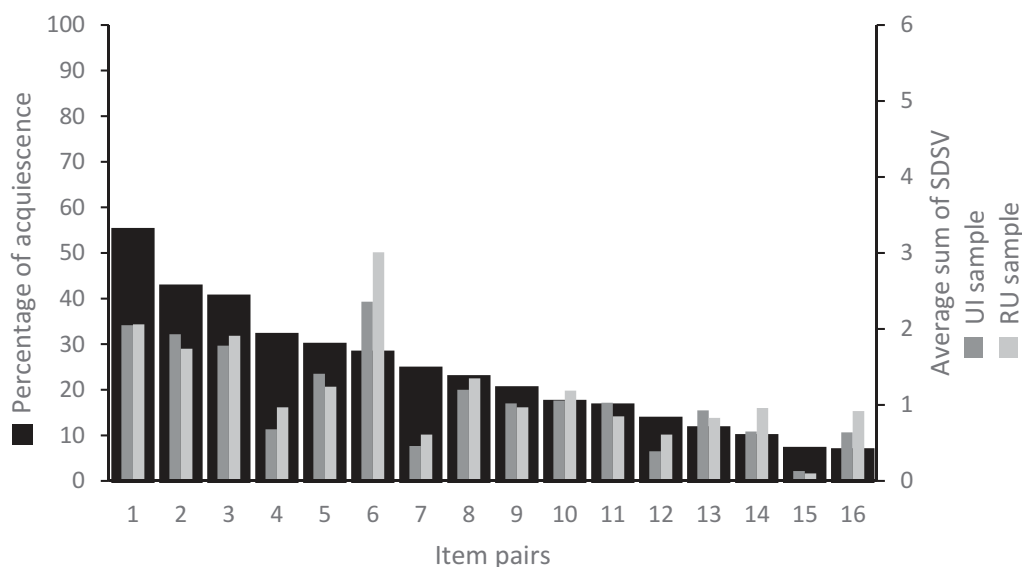
The mean response to the set of the 16 item-pairs is recommended as an index of acquiescence, as agreeing with both items in an item-pair is taken to represent acquiescent responding. In the current study, however, acquiescence was estimated by, first, recoding each of the 32 items in such a way that both 'Agree' and 'Agree strongly' were coded as 1 and other responses ('Disagree strongly', 'Disagree' and 'Neither agree nor disagree') were coded as zero. Next, the items in each item-pair were summed up and the percentage of respondents agreeing with both items (sum score of 2) calculated for each item-pair.

**2.1.2.3. Social Desirability Scale Value (SDSV).** The SDSV of the BFI items was evaluated by asking respondents to rate the desirability of each item's content. The question: 'How desirable or undesirable is it to possess the following characteristics?' was presented on a seven-point scale with the response options: 'Very desirable', 'Rather desirable', 'A little desirable', 'Neither desirable nor undesirable', 'A little undesirable', 'Rather undesirable' and 'Very undesirable'. Responses were coded in such a way that higher scores indicated more desirable content (coded from 3 to -3). The measure of SDSV for each item-pair was computed by summing up SDSV scores on both items and calculating the mean of the sum scores for each item-pair, producing a mean rating of the characteristic as expressed by each of the two items making up the item-pair.

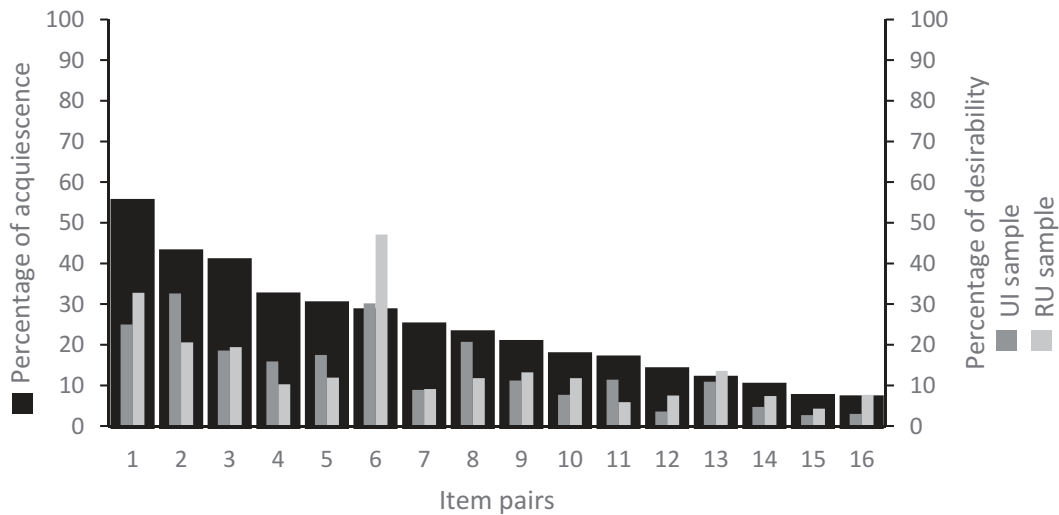
## 2.2. Results and discussion

The results showed that percentages of acquiescence varied considerably between the item-pairs and that the pattern of percentages of acquiescence strongly resembles the patterns of average SDSVs for each item-pair as can be seen in Figure 1, where item-pairs have been ordered from the pair with the highest percentage of acquiescence to the pair with the lowest percentage. Furthermore, the correlation between the percentages of respondents agreeing with both items in an item pair and the average SDSV for the item pair was  $r = .754$ ,  $p = .001$  for SDSV's from the UI sample and  $r = .658$ ,  $p = .006$  for SDSV's from the RU sample. The results thus suggest that the measure of acquiescence is inflated by the desirability of item content in an item pair, which can be taken as evidence for the hypothesized effect of socially desirable responding upon the measure of acquiescence.

Another way in which the two patterns can be compared is to compare the percentages of respondents agreeing with both items in an item pair with the percentages of respondents rating



**Figure 1.** Valid percentage of respondents agreeing with both items in each item pair (in black) and mean sum of SDSV for item pairs (in grey).



**Figure 2.** Comparison of percentages of respondents agreeing with both items in each item-pair (in black) with percentages of respondents rating both items as desirable in each item-pair (in grey).

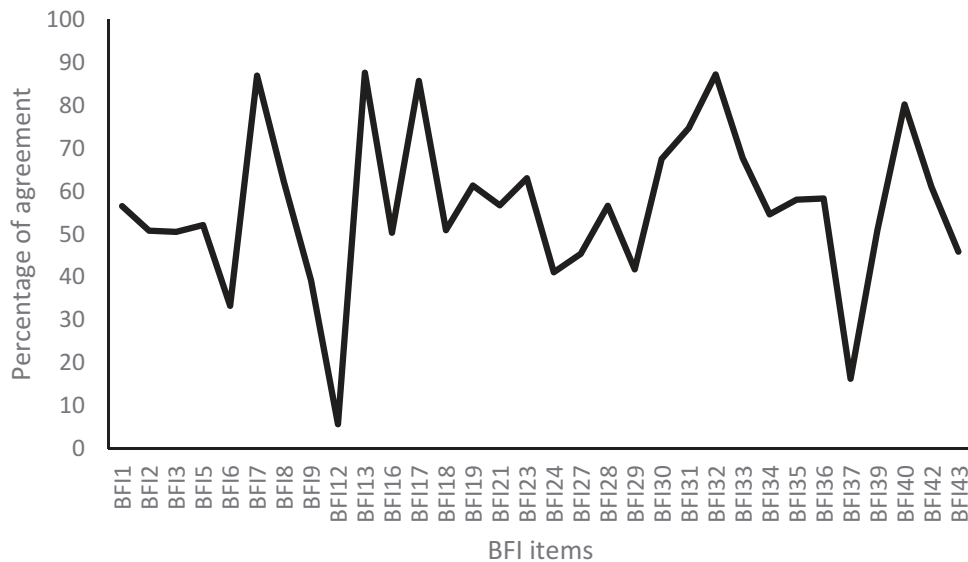
both items as desirable (percentage of respondents selecting either ‘Very desirable’, ‘Rather desirable’ or ‘A little desirable’ on both items). As can be seen from Figure 2 the patterns are very similar and the measures are highly correlated in the UI sample ( $r = .809, p < .001$ ) and in the RU sample ( $r = .613, p = .012$ ).

This comparison, however, raises the question of an alternative interpretation of the findings, namely that the measure of SDSV was affected by acquiescence (in such a way that rating an item as desirable is a form of agreement) and therefore produced a similar pattern of percentages of respondents rating both items in an item-pair as desirable to the pattern of percentages of respondents agreeing with both items. This interpretation of the results would call for an explanation of the observed pattern in acquiescent responding. As discussed earlier, lack of cognitive processing can result in acquiescent responding (see e.g. Gilbert (1991)). There are several factors which can limit cognitive processes, however, only factors that would produce the observed pattern apply here. Therefore, individual differences in cognitive processes need no further discussion in this context as differences in acquiescence between item-pairs cannot be explained by individual variability in the tendency to agree to items regardless of the content.

Turning to situational factors, the same logic applies to external distracting factors as individual differences because the data was collected online and therefore such distractions vary between individuals and thus cannot be expected to result in any specific pattern in the data. Therefore, if this is the expected pattern of acquiescence in the BFI items, regardless of instructions and response scales, this pattern must be brought about by some characteristics of the items forming the item-pairs.

A characteristic of the BFI (as most other psychological questionnaires) is that the items are presented in a specific order. If item order affects acquiescence, it is reasonable to presume that items presented later in questionnaires would be more prone to acquiescent responding due to decreasing motivation and/or increasing fatigue, which would reduce cognitive processing. This could produce a specific pattern in item-pair agreement that would reflect the position of the items forming each item-pair, i.e. item agreement would increase as the item appears later in the questionnaire and item-pair agreement would therefore depend on the placement of the two items within the questionnaire. However, as can be seen from Figure 3, item agreement does not increase as the item is presented later in the questionnaire and therefore item order does not seem to affect to item agreement.

Another characteristic of items thought to increase acquiescence is item difficulty (see e.g. Krosnick, 1991). Study II was thus conducted to evaluate the difficulty of the items forming the item pairs.



**Figure 3.** Percentages of agreement with BFI items belonging to item-pairs in presentation order.

### 3. Study II: evaluation of item difficulty

Study II was aimed at evaluating item difficulty of the 32 BFI items from the 16 item-pairs with Cognitive Interviews (CogI). The interviews were conducted on a sample of 12 interviewees and focused on content understanding.

#### 3.1. Method

##### 3.1.1. Participants and procedure

Cognitive interviews were conducted on a convenience sample of 12 participants, eight females and four males, between 19 and 69 years of age. The interviews were semi-structured with probing questions on each term within each of the 32 BFI items forming the item-pairs (see Willis, 2005 for more information on CogI).

#### 3.2. Results

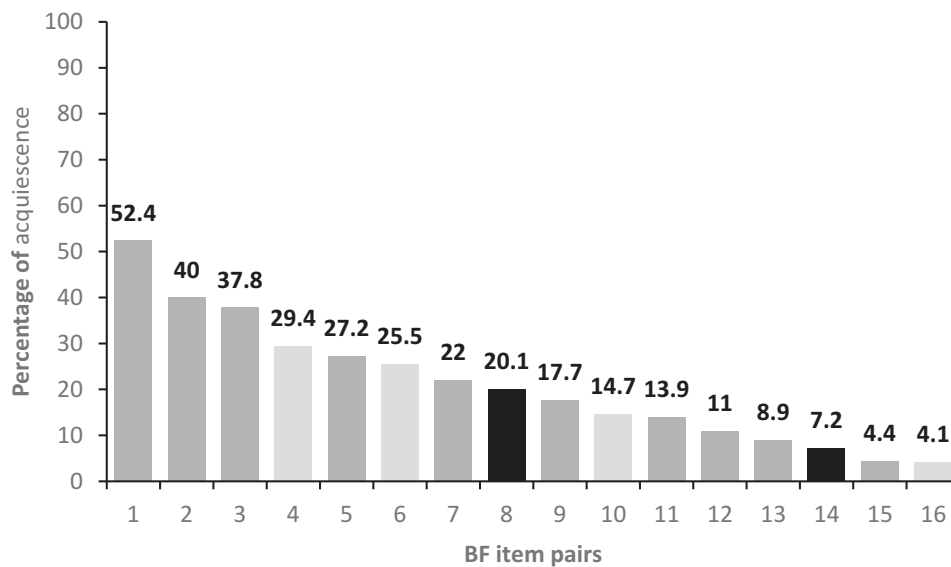
In the cognitive interviews, indicators of item difficulty were observed in 14 out of the 32 items comprising the item-pairs. Indicators of item difficulty were classified as follows: Understanding of words (U): Items in this category are items where one or more participant found it difficult to understand a word used in the item; Double barreled items (D): Items with more than one descriptive term which participants found to have different implications for personality; Response difficulty (R): Items that participants found it difficult to respond to even though they understood the content of the item. The findings from the cognitive interviews are shown in Table 1.

As can be seen from Table 1 the percentages of respondents agreeing with an item vary between the items classified under one or more of the difficulty categories and are on average not higher (mean = 55.1%,  $SD = 21.2\%$ ), than percentages of those agreeing with the items that did not fall under any difficulty category (mean = 54.0%,  $SD = 17.7\%$ ),  $t(30) = .163$ ,  $p = .872$ . Therefore, there is no indication of respondents agreeing more often with items identified as problematic in the cognitive interviews. Taking a look at the percentages agreeing with both items in an item-pair in Study I, the findings from the cognitive interviews do not correspond with item-pair agreement. This can be seen in Figure 4, where the bars representing the percentages agreeing with both items in an item-pair have been colored to represent an indication of difficulty in neither item (light grey), in one item (dark grey) and both items (black).

**Table 1.** Indicators of item difficulty identified with cognitive interviews, percentages of respondents agreeing with the item in study I and percentages agreeing with both items in the item-pair to which the item belongs in study I.

Pair	Item	Cogl	% Item agree	% Pair agree
1	13	U	87.6	52.4
2	17	U	85.7	40.0
3	36	D	58.3	37.8
5	33	U	67.7	27.2
7	27	U	45.4	22.0
8	1	U	57.0	20.1
8	21	U	56.7	17.7
9	34	R	54.6	17.7
11	9	D	38.9	13.9
12	37	U	16.3	11.0
13	6	U	33.6	8.9
14	24	DR	41.1	7.2
14	29	U	41.8	7.2
15	7	D	86.8	4.4

Cogl: U: Understanding of words; D: Double barreled item; R: Response difficulty.



**Figure 4.** Percentages of respondents agreeing with both items in an item-pair with the bars colored to represent difficulty in neither item (light grey), in one item (dark grey) and in both items (black).

#### 4. General discussion and conclusion

The results of Study I indicated that the 16 item BFI measure of acquiescence was affected by socially desirable responding. Study II provided further support for this finding, as the items identified as difficult in the cognitive interviews were not endorsed more frequently than other items, contrary to what would have been expected if the underlying response mechanism was due to acquiescent responding. The overall finding is thus that item difficulty (as measured with cognitive interviews) did not produce more agreement and therefore the findings indicate that socially desirable responding was largely responsible for the observed agreement in the item-pairs selected by Soto et al. (2008) to measure acquiescence.

These results call for further evaluation of measures of acquiescence, focusing on the item content and especially the desirability of the content. As explained in the introduction to this paper, social desirability scales are aimed at measuring a person's general tendency to respond to in a socially desirable manner but should not be expected to correlate with agreement in general. It is therefore important for research on acquiescence and socially desirable responding to focus on item level evaluation of item-pairs presumed to be comprised of items with opposite implications.

There are two obvious limitations to this research. The first is the use of convenience samples and the second is the use of the Icelandic version of the BFI. Both of these factors limit the generalizability of the findings and call for further evaluation of the reported effect. Future research could also be directed at response times for the BFI items as socially desirable responding is presumed to occur after considerable cognitive processing of an item and should thus produce much longer response times than acquiescent responding, which is thought to be the result of insufficient cognitive processing. In conclusion, the assessment of causes of response biases with a focus on cognitive processes is a highly important topic that needs to be researched further.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

**Vaka Vésteinsdóttir** is an adjunct professor of Psychology, at the University of Iceland, School of Health Sciences, project manager at the School of Health Sciences' Testing Center and director of RAHÍ Methodology Research Center. She obtained her BA and MA degrees at the University of Iceland on the topics of telephone interviewers' voice and acceptance rates, and honesty testing. After receiving her PhD in psychological methodology from the University of Iceland she was postdoc in the Psychological Methods, Assessment and iScience team at the University of Konstanz. Her research interests are in survey methodology, particularly in questionnaire construction and measurement bias, with a focus on socially desirable responding.

**Ragnhildur Lilja Asgeirsdóttir** is a PhD student in Quantitative Psychology at the University of Iceland, Department of Psychology. She obtained her Bachelor and Masters degrees at the University of Iceland. Her research interests are in the areas of survey methodology and psychometrics, particularly in questionnaire construction and measurement bias.

**Fanney Thorsdóttir** is an associate professor of Psychology at the University of Iceland, School of Health Sciences and co-founder of RAHÍ Methodology Research Center. She has a BA and a Master (Psychology) from the University of Iceland, a Diploma in Social Science Data Analysis from the University of Essex and a PhD (psychology) from the London School of Economics. Her research interests are in the areas of psychometrics and survey methodology with a particular focus on measurement bias in self-report measurements.

**Ulf-Dietrich Reips** is a full professor in the Faculty of Sciences at the University of Konstanz, where he holds the Chair for Psychological Methods, Assessment, and iScience. For more than two decades he has been working on Internet-based research methodologies (or Internet science), the psychology of the Internet, measurement, development, the cognition of causality, personality, privacy, Social Media, crowdsourcing, and Big Data. In 1994, he founded the Web Experimental Psychology Lab, the first laboratory for conducting real experiments on the World Wide Web. Ulf was elected the first non-North American president of the Society for Computers in Psychology (SCiP) and he is the founding editor of the International Journal of Internet Science. Many of his over 140 publications are among the most highly cited in their journals, see <http://www.uni-konstanz.de/iscience/reips/pubs/publications.html>. Ulf has worked, lived, and studied in California, Colorado, Israel, Germany, Spain, Switzerland, and the UK. In 2014, he was ranked 7th of 'Top Scientists working at Spanish Private Universities' by the Consejo Superior de Investigaciones Científicas, Spain. Recently, he has been asked to direct the Leibniz institute for Psychology information in Trier, Germany. Ulf and his team develop and provide free Web tools for researchers, teachers, students, and the public. They received numerous awards for their Web applications (available from the iScience Server at <http://iscience.eu/>) and methodological work serving the research community.

## References

- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *The Journal of Abnormal and Social Psychology*, 60(2), 151–174. <https://doi.org/10.1037/h0040372>
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6(4), 475–494. <https://doi.org/10.1177/001316444600600405>
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Dryden.



- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*(2), 155–166. <https://doi.org/10.1037/0021-9010.84.2.155>
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*(2), 107–119. <https://doi.org/10.1037/0003-066X.46.2.107>
- Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science, 39*(3), 184–201. <https://doi.org/10.1037/cjbs2007015>
- John, O. P., Donahue, E., & Kentle, R. (1991). ‘The “big five. *Inventory—version 4a and, 54.*
- Knowles, E. S., & Nathan, K. T. (1997). Acquiescent responding in self-reports: Cognitive style or social concern? *Journal of Research in Personality, 31*(2), 293–301. <https://doi.org/10.1006/jrpe.1997.2180>
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236. <https://doi.org/10.1002/acp.2350050305>
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). Academic Press.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718–737. <https://doi.org/10.1037/0022-3514.94.4.718>
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*(3), 299–314. <https://doi.org/10.1037/0033-2909.103.3.299>
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality assessment. *Educational and Psychological Measurement, 59*(2), 197–210. <https://doi.org/10.1177/00131649921969802>
- Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research, 49*(5), 737–747. <https://doi.org/10.1509/jmr.11.0368>
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage.