Do simpler item wording and response scales reduce acquiescence in personality inventories? A survey experiment

Rammstedt, Beatrice; Roemer, Lena; Lechner, Clemens

Contents lists available at ScienceDirect

Personality and Individual Differences

Short Communication

Do simpler item wording and response scales reduce acquiescence in personality inventories? A survey experiment

Beatrice Rammstedt [*], Lena Roemer, Clemens M. Lechner

GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany

A B S T R A C T

Acquiescence ("yea-saying") presents a major challenge in personality assessment via questionnaire. Cognitive load when responding to items seems to be related to a greater tendency to acquiesce. The aim of this study was to experimentally investigate whether reducing the complexity of personality items through simplified item wording and response scales with fewer options reduces acquiesce. Based on a heterogeneous sample of almost 6000 respondents, we probed acquiescent responding on the 60-item Big Five Inventory-2 (BFI-2) in a 2 × 2 design, comparing (a) the original items and simplified versions and (b) the original 5-point and a 4-point response scale without a midpoint. Results show that the response scale format did not affect acquiescent responding. Contrariwise, item simplification reduced the variance attributable to acquiescence in the total sample. Moderation analysis revealed stronger effects for lower-educated respondents.

1. Introduction

Response styles present a major challenge when assessing personality constructs. Among the most frequently observed response styles in self-report surveys is acquiescence, an individual's tendency to agree with survey items regardless of their content (Jackson & Messick, 1958). Acquiescence can introduce substantial bias into individuals' item responses, jeopardizing the construct validity of personality instruments and undermining researchers' ability to draw valid inferences concerning the relationship of personality to other constructs of interest (Van Vaerenbergh & Thomas, 2013). The question as to what factors might account for acquiescent responding in personality assessment is thus an important one.

Previous—mostly correlational—studies have established several correlates and determinants of acquiescent responding (for an overview and theoretical framework, see Lechner, Partsch, Danner, & Rammstedt, 2019). Most have focused on respondent characteristics and identified inter alia lower education (e.g., Rammstedt, Goldberg, & Borg, 2010) and/or cognitive ability (Gudjonsson, 1986; Lechner & Rammstedt, 2015) as the most potent and robust determinants of acquiescent responding. Others have examined how questionnaire characteristics might influence acquiescent responding and have identified, in particular, the cognitive load of the items as a factor that increases acquiescent responding (Knowles & Condon, 1999). The fact that understanding

ambiguously and complexly formulated items requires greater cognitive effort (Lenzner, Kaczmirek, & Galesic, 2011) gives reason to assume that the cognitive-processing capacity available for answering items is a major determinant of acquiescence (Lechner, Partsch, Danner, & Rammstedt, 2019). If items are lengthy and/or complex (Condon, Ferrando, & Demestre, 2006) and ambiguous (Angleitner, John, & Löhr, 1986), they require greater cognitive processing, which might induce stronger acquiescence, especially in respondents whose cognitive processing capacity is generally lower (as indexed, e.g., by lower education and/or cognitive ability; Lechner & Rammstedt, 2015). Previous research (Condon, Ferrando, & Demestre, 2006) has indicated that lengthier and more complex items were more susceptible to acquiescence. Therefore, one can assume that reducing the complexity and ambiguity of personality items will reduce the tendency for acquiescent responding for these items —especially among respondents whose cognitive processing capacity is generally lower.

The aim of the present study was to experimentally investigate whether reducing the complexity of personality items reduces acquiescent responding. We hypothesized, first, that using simplified item wording and a simplified response scale would reduce acquiescent responding in terms of the variance attributable to acquiescence (rather than the substantive personality factors); and, second, that these effects would be more pronounced in lower-educated respondents.

To test our hypotheses, we experimentally compared different

---

versions of the Big Five Inventory-2 (BFI-2; Soto & John, 2017). As most of the original BFI-2 items are either lengthy or double-barreled (i.e., contain multiple stimuli; see Schult, Schneider, & Sparfeldt, 2019), they can be considered complex by established item-formulation standards (see also Rammstedt, Roemer, Danner, & Lechner, 2022). In addition, the 5-point response scale used has been criticized as too complex for weaker readers and/or lower-educated respondents (e.g., Borgers, Hox, & Sikkel, 2004). Further, some researchers (e.g., Weijters, Cabooter, & Schillewaert, 2010) have argued that including a mid-point (as in a 5-point scale) increases acquiescent responding. In a 2 × 2 experimental design, we therefore compared the original BFI-2 items and its 5-point response scale with (a) simplified versions of the items optimized for readability and (b) a 4-point response scale. We tested these four conditions in a large, heterogeneous sample rather than in the usual sample of psychology students, who are generally less prone to acquiescent responding.

2. Method

2.1. Sample

Data were collected in summer 2016 via the Survey Monkey platform as part of the Programme for the Assessment of Adult Competencies (PIAAC) English Pilot Study on Non-Cognitive Skills (Organisation for Economic Co-operation and Development (OECD), 2018). Participants aged 16–65 years were selected in the USA and the UK according to a quota scheme based on sex, age, and region, broadly representative of U. S. and UK census data.

From the total sample, around 25 % of respondents were excluded by the OECD because they failed more than one of eight quality checks (see OECD, 2018).[1] This resulted in a quality-controlled sample of $N = 5910$ respondents (57 % female), of whom $n = 1481$ (25 %) were classified as lower-educated (high school diploma or lower) based on their self-reported highest level of educational attainment.

2.2. Instruments: original and simplified BFI-2

The BFI-2 (Soto & John, 2017) assesses the Big Five personality domains and three central facets per domain. It comprises 60 phrase-like items of which about half contain separate, albeit similar, stimuli and thus can be regarded as double-barreled (see Schult, Schneider, & Sparfeldt, 2019). The Flesch Reading Ease score was $Mdn = 59.75$ ($SD = 41.94$), suggesting that the items are on average fairly difficult (for details, see Rammstedt, Roemer, Danner, & Lechner, 2022). Therefore, the expert group responsible for developing the PIAAC Pilot[2] decided to reduce the complexity of the original BFI-2 in order to make it more appropriate for use with the general adult population—especially the less literate members (OECD, 2018, p. 1). Following this guiding principle, they developed simplified versions of 55 of the 60 items (for details of the simplification process, see Rammstedt, Roemer, Danner, & Lechner, 2022), resulting in items with a Flesch Reading Ease score of $Mdn = 76.89$ (SD = 41.20), which indicates fairly high readability on average.

2.2.1. Response format

The BFI-2 items are typically answered using a 5-point rating scale with the options strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), and strongly agree (5). As the expert group also considered this response format difficult (especially given the neutral midpoint), they tested a 4-point version without the midpoint (OECD, 2018).

2.3. Experimental design

The original and simplified item versions and the two response scales were randomly assigned to four experimental conditions (2 × 2). In Condition A ($n = 1442$), respondents received the original item wording and the original 5-point response scale; in Condition B ($n = 1416$), the simplified items and the 5-point response scale were administered; in Condition C ($n = 1521$), the original items were combined with the 4-point scale; and in Condition D ($n = 1531$), the simplified items and the 4-point scale were used.

2.4. Estimating acquiescence

Because the BFI-2 comprises a balanced set of positively and negatively keyed items (see Soto & John, 2017), individual differences in acquiescent responding can be explicitly modeled through an acquiescence factor as suggested by Billiet and McClendon (2000).[3] The Billiet and McClendon (2000) CFA approach and its various extensions, such as random-intercept EFA (Aichholzer, 2014) or the domains-incremental-facets-acquiescence-bifactor model (DFAB; Danner, Lechner, Soto, & John, 2021), are a widely used, de-facto standard approach to modeling acquiescence in personality research. It has been shown to perform well and to be robust to violations of various assumptions (Savalei & Falk, 2014). This approach was ideally suited for our present study, whose intent was to separate the variance attributable to acquiescence from the construct variance in the Big Five. Therefore, we ran a series of confirmatory factor analyses. The models included five latent factors representing the Big Five domains and one acquiescence method factor, whereas the five domain factors were allowed to correlate, the acquiescence factor was orthogonal. Modeling the acquiescence factor as orthogonal is in line with the definition of acquiescence as a tendency to agree with items independent of their content and keying, and is also necessary for the model to be estimable from the same item set as the substantive factors.[4]

To be able to model the five domains simultaneously, we used six item parcels as indicators for each domain factor, each containing two same-keyed items for the facet in question. Loadings on the acquiescence factor were constrained to 1.[5] Our main motivation for using item parcels compared to single items were (1) to reduce model complexity and circumvent convergence issues, and (2) to obtain input variables for the models that have better distributional properties. Different from the standard approach of the BFI-2 (Soto & John, 2017), we did not use the personality facet scores (i.e., parcels) as indicators for the domain factors; this approach was not viable for our present study because the facet scores are balanced-keyed, consisting of an equal number of positively and negatively worded items. As such, they are implicitly controlled for acquiescence. This makes it impossible to model an acquiescence factor

---

[1] These quality checks applied by the OECD included country of residence, age, testing time, ability test, consistency of answers, and answers on quality control questions like "I fly to the International Space Station".

[2] The members of the expert group were Daniel Danner, Beatrice Rammstedt, Brent Roberts, Richard Roberts, Manfred Schmitt, Fons van de Vijver, and Susanne Weiß.

[3] An alternative method to statistically control for acquiescence in a balanced set of items is ipsatization, thus to use mean-corrected "deviation scores" instead of the original raw scores. As we, however, changed the item formulations between our conditions, differences in these mean-corrected scores could not unambiguously be ascribed to differences in acquiescent responding.

[4] This does not imply that acquiescence is uncorrelated with other respondent characteristics such as cognitive ability or deferential communication styles. However, such correlations need to be estimated from different item sets than the one through which acquiescence is identified (e.g., Lechner, Partsch, Danner, & Rammstedt, 2019).

[5] To test that the BFI-2 responses were in fact affected by acquiescence, we tested our solution against a model without the acquiescence factor. For each experimental condition, likelihood ratio tests indicated worse fit when the acquiescence factor was omitted.

in the way that was required for our study.

The parameter of interest compared across the four conditions was the variance of the acquiescence factor, which indicates the amount of variance in the items that arises from the systematic tendency to agree (or disagree) with the items, rather than from the substantive Big Five factors. The higher the variance of the acquiescence factor, the more pronounced are individual differences in the tendency to acquiesce—and the greater the distortion of factor structures, correlations, and means potentially becomes if acquiescence would be unaccounted for.

As it could be argued that the amount of acquiescence variance is not directly comparable in the two conditions with a 4-point scale versus the two conditions with a 5-point scale, we additionally used an alternative procedure for estimating the tendency for acquiescence in these conditions. Specifically, for each condition, we computed the ratio of the acquiescence variance to the total variance (i.e., the sum of the acquiescence variance, Big Five factor variance, and error variance). These ratios are not differentially affected by the different response scale formats.

3. Results

We hypothesized, first, that the two forms of simplification—namely, simplifying the item wording and the response scale—would reduce the tendency to acquiesce; and, second, that this reduction would be more pronounced in lower-educated respondents. To test these hypotheses, we estimated—separately for the four experimental conditions—the variance of the acquiescence factor in the models described above. We then tested differences in the variance across the four experimental conditions (e.g., original item wording, 5-point response scale vs. simplified item wording, 5-point response scale) with F tests; those for the ratios with Wald-tests. To account for multiple testing, we used Holm–Bonferroni correction and interpreted adjusted p values $\leq .01$ as significant. The variance of the acquiescence factor across the four experimental conditions both for the total sample and for high- versus low-educated respondents are displayed in Fig. 1; more detailed results, and further information such as variances for the domain factors or correlations among scale scores, are provided on OSF.[6] Results for the ratio scores computed as an alternative test for differences in acquiescence across the different response scale formats are shown in Fig. 2.

Results reveal a clear main effect for wording. Variance of the acquiescence factor was systematically lower in the two conditions with simplified wording (B, D) compared with those with the original wording (A, C). Variance due to acquiescence was 3.9 % in Condition A vs 3.0 % in Condition B, $F(1441, 1415) = 1.32, p < .001$; it was 4.1 % in Condition C vs 3.6 % in Condition D, $F(1520, 1530) = 1.14, p = .01$. By contrast, based on both forms of acquiescence computation, no effect was found for the simplification of the response scale (all $p > .01$). Comparison of the ratios largely replicated the pattern of the acquiescence-reducing effect of simplifying the wording and no main effect of simplifying the response scale (p = n.s.). Thus, our first hypothesis is supported with regard to wording simplification but not response scale format.

When we split our sample into higher- and lower-educated respondents, the acquiescence-reducing effect of simplified item wording found for the total sample was—as hypothesized—even more pronounced for lower-educated respondents. In the two conditions with simplified item wording (B, D), the acquiescence factor exhibited substantially less variance than in the conditions with the original wording (A, C): A = 4 % vs B = 2.2 %, $F(394, 355) = 1.81, p < .001$; C = 4.6 % vs D = 3.5 %, $F(360, 368) = 1.33, p < .01$. For higher-educated respondents—who have been found to be generally less prone to acquiescent responding—a more homogeneous picture emerged. Only in the 5-point

response scale conditions (A, B) did these respondents show on average a slightly lower tendency for acquiescent responding: A = 3.9 % vs B = 3.2 %, $F(1046, 1059) = 1.21, p < .01$; C = 3.9 % vs D = 3.6 %, $F(1159, 1161) = 1.08, p = .09$. Again, simplifying the response scale did not systematically reduce acquiescent responding.

4. Discussion

Does reducing the linguistic complexity of personality items reduce respondents' tendency to acquiesce? We investigated this question by experimentally testing the effects of simplifying the item wording and the response scale on the tendency for acquiescent responding in the total sample and in subsamples comprising higher- and lower-educated respondents. Our study supports and extends previous research by showing that (a) simplifying item wording generally reduced acquiescent responding and (b) this effect was even more pronounced in lower-educated respondents. By contrast, our results contradicted previous findings (e.g., Weijters, Cabooter, & Schillewaert, 2010) in that we found (c) that omitting the midpoint in the response scale did not affect acquiescent responding.

Our study design was limited in different ways. First, the exclusion of participants that failed several quality checks (e.g., attention checks) likely increased the response quality in the remaining sample. By that respondents with a high tendency for acquiescence might have been deleted from the sample. Second, double-barrelledness of items can be regarded as a continuum from very similar stimuli to stimuli that are even contradictory (see Rammstedt, Roemer, Danner, & Lechner, 2022). In the case of BFI-2, the items are clearly on the synonym pole of this dimension and are thus probably less confusing than items containing contradictory stimuli. Both aspects might have led to a generally reduced acquiescence in the present study. Third and finally, in our present study we intentionally tested the effect of an inclusion or omission of a response scale midpoint. We did not investigate the general effect of a simple response format (e.g., only binary) versus progressively increasing response categories on acquiescence – a design that should be addressed in future studies.

In sum, our study shows that using simpler items (but not different response formats) can reduce acquiescence, especially in lower-educated individuals. Of note, this does not necessarily imply that simpler items are universally preferable: As Rammstedt, Roemer, Danner, and Lechner (2022), using the same data, found, simplified items also entail a lower reliability and validity of the scale scores, likely because the simpler items were also less precise or informative than the original items. Thus, the present findings on the reducing effects on acquiescence, especially among those with lower education, highlight the importance of a thorough item development with a specific notion to non-complex item phrasings; while the earlier findings by Rammstedt, Roemer, Danner, and Lechner (2022) suggest that item developers have to strike a balance between simplicity and information density if they want to obtain good reliability and validity. In our view, the beneficial effect of simplified items in terms of reducing acquiescence found in the present study is still outweighed by the lower information density of the simplified version, which jeopardizes reliability and validity. Yet, the answer of what item wording is simple what is not-yet-too-simple may depend on the specific application and target population.

CRediT authorship contribution statement

Beatrice Rammstedt: Conceptualization, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Lena Roemer:** Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Clemens M. Lechner:** Methodology, Writing – original draft, Writing – review & editing.
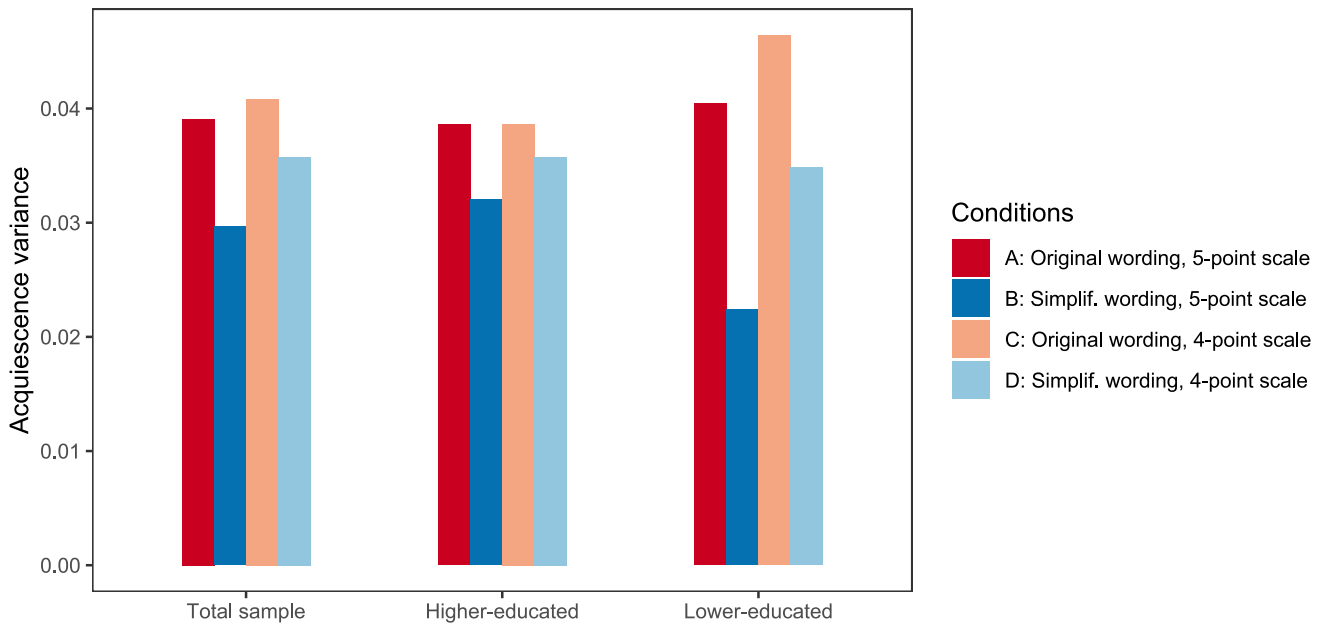

Fig. 1. Variance of the acquiescence factor in the four experimental conditions and separately for the total sample and for the higher-educated and lower-educated subsamples.
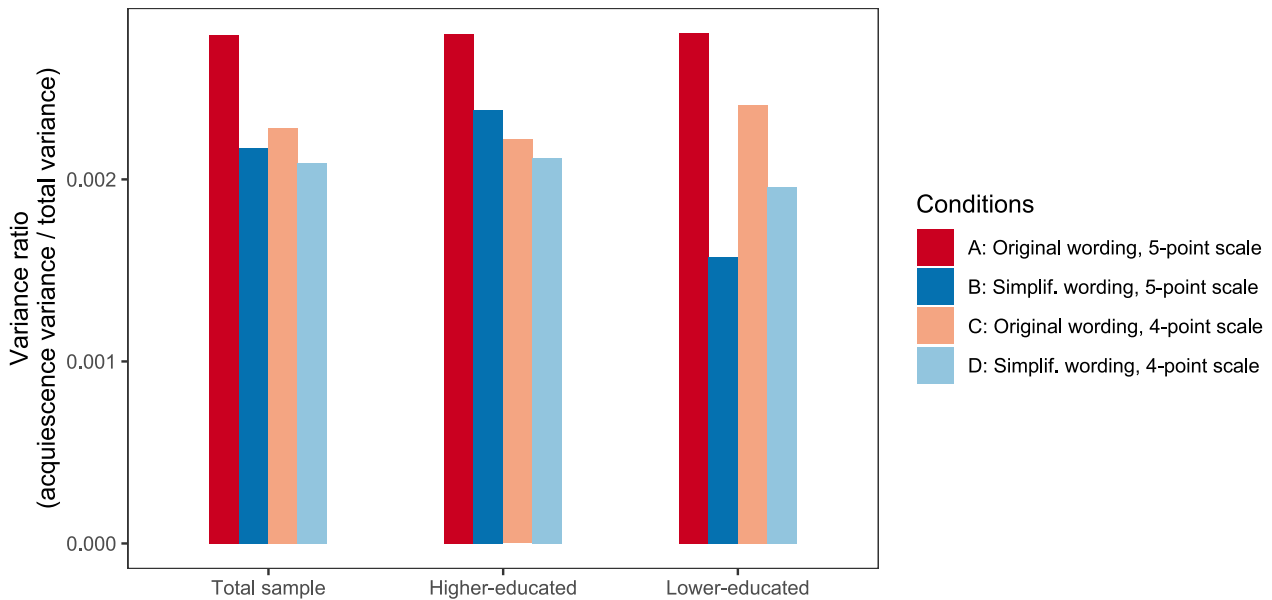

Fig. 2. Ratio of the acquiescence variance to the total variance in the four experimental conditions and separately for the total sample and for the higher-educated and lower-educated subsamples.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is available through a research data center. Information is provided.

References

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality, 53*, 1–4. https://doi.org/10.1016/j.jrp.2014.07.001

Angleitner, A., John, O. P., & Löhr, F.-J. (1986). It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner, & J. S. Wiggins (Eds.), *Personality assessment via questionnaires. Current issues in theory and measurement* (pp. 61–107). Berlin: Springer.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*(4), 608–628. https://doi.org/10.1207/S15328007SEM0704_5

Borgers, N., Hox, J., & Sikkel, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality & Quantity, 38*(1), 17–33. https://doi.org/10.1023/B:QUQU.0000013236.29205.a6

Condon, L., Ferrando, P. J., & Demestre, J. (2006). A note on some item characteristics related to acquiescent responding. *Personality and Individual Differences, 40*(3), 403–407.

Danner, D., Lechner, C. M., Soto, C. J., & John, O. P. (2021). Modelling the incremental value of personality facets: The Domains-Facets-Acquiescence-Bifactor (DFAB) model. *European Journal of Personality, 35*, 67–87. https://doi.org/10.1002/per.2268

Gudjonsson, G. H. (1986). The relationship between interrogative suggestibility and acquiescence: Empirical findings and theoretical implications. *Personality and Individual Differences, 7*(2), 195–199. https://doi.org/10.1016/0191-8869(86)90055-3

Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin, 55*(4), 243–252. https://doi.org/10.1037/h0045996

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77*(2), 379–386. https://doi.org/10.1037/0022-3514.77.2.379

Lechner, C. M., Partsch, M. V., Danner, D., & Rammstedt, B. (2019). Individual, situational, and cultural correlates of acquiescent responding: Towards a unified conceptual framework. *British Journal of Mathematical and Statistical Psychology, 72*(3), 426–446. https://doi.org/10.1111/bmsp.12164

Lechner, C. M., & Rammstedt, B. (2015). Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychological Assessment, 27*(4), 1301–1311. https://doi.org/10.1037/pas0000151

Lenzner, T., Kaczmirek, L., & Galesic, M. (2011). Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of Public Opinion Research, 23*(3), 361–373. https://doi.org/10.1093/ijpor/edq053

Organisation for Economic Co-operation and Development (OECD). (2018). *Programme for the International Assessment of Adult Competencies (PIAAC), English Pilot Study on Non-Cognitive Skills (ZA6940; Version 1.0.0) GESIS Data Archive.* https://doi.org/10.4232/1.13062

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of big-five factor markers for persons with different levels of education. *Journal of Research in Personality, 44*(1), 53–61. https://doi.org/10.1016/j.jrp.2009.10.005

Rammstedt, B., Roemer, L., Danner, D., & Lechner, C. M. (2022). Don't keep it too simple: Simplified items do not improve measurement quality. *European Journal of Psychological Assessment.* https://doi.org/10.1027/1015-5759/a000741. Advance online publication.

Savalei, V., & Falk, C. F. (2014). Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate Behavioral Research, 49*, 407–424. https://doi.org/10.1080/00273171.2014.931800

Schult, J., Schneider, R., & Sparfeldt, J. (2019). Assessing personality with multi-descriptor items: More harm than good? *European Journal of Psychological Assessment, 35*(1), 117–125. https://doi.org/10.1027/1015-5759/a000368

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*, 117–143. https://doi.org/10.1037/pspp0000096

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research, 25*(2), 195–217. https://doi.org/10.1093/ijpor/eds021

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27*(3), 236–247. https://doi.org/10.1016/j.ijresmar.2010.02.004