MONASH University

MONASH BUSINESS SCHOOL

Department of Economics
ISSN number 1441-5429

Improving the reliability and validity of data on Big Five personality traits in developing countries

Discussion Paper no. 2022-04

An Huang and Paulo Santos

Abstract:

Recent research argues that personality data collected in developing countries may fail to measure the intended personality traits. We quantify the importance of three potential drivers of such failure (acquiescence response, social desirability response, and enumerator effects) in terms of their influence on psychometric properties and the predictive power of data on the Big Five personality traits. Our results suggest that both the reliability and validity of survey data on the Big Five can be improved by correcting for acquiescence bias, and that the performance of this correction can be greatly increased through the use of balanced personality inventories. Correcting for social desirability response and enumerator effects matters when estimating the effect of personality traits on economic outcomes (but not when capturing personality structure).

Keywords: balanced questionnaire, personality traits, acquiescence bias, social desirability bias

JEL Classification: B49, C83, D91

An Huang: Economics, Monash University (email: an.huang@monash.edu); Paulo Santos: Economics, Monash University (email: paulo.santos@monash.edu).

Improving the reliability and validity of data on Big Five personality traits in developing countries[1]

An Huang[*†‡], **Paulo Santos**[*‡]

Abstract: Recent research argues that personality data collected in developing countries may fail to measure the intended personality traits. We quantify the importance of three potential drivers of such failure (acquiescence response, social desirability response, and enumerator effects) in terms of their influence on psychometric properties and the predictive power of data on the Big Five personality traits. Our results suggest that both the reliability and validity of survey data on the Big Five can be improved by correcting for acquiescence bias, and that the performance of this correction can be greatly increased through the use of balanced personality inventories. Correcting for social desirability response and enumerator effects matters when estimating the effect of personality traits on economic outcomes (but not when capturing personality structure).

Keywords: balanced questionnaire, personality traits, acquiescence response, social desirability response.

[1] We thank M. Brent Donnellan, Rachid Laajaj and their coauthors for making the original data and code from their studies accessible for this work. Paulo Santos acknowledges funding support from the Australian Centre for International Agricultural Research (ACIAR) Grant ASEM/2012/073. All data and code are available for review upon request, and they will be made publicly available upon publication. The authors declare no competing interests.

[*] Department of Economics, Monash University, 900 Dandenong Road, 3145 Caulfield, Victoria, Australia.

[†] Corresponding author. Email address: an.huang@monash.edu.

[‡] These authors contributed equally to this work.

1. Introduction

Over the past decades, economists have paid increasing attention to the importance of personality traits as predictors of important life outcomes, including success in education (Burks et al., 2015), job performance and earnings (Donato, Miller, Mohanan, Truskinovsky, & Vera-Hernández, 2017; Gensowski, 2018); marriage (Dupuy & Galichon, 2014), health outcomes (Savelyev & Tan, 2019); and cooperative and pro-social behavior (Drouvelis & Georgantzis, 2019; Proto, Rustichini, & Sofianos, 2019)[2]. Most of the analysis focused on the role of the Big Five personality traits (i.e., Agreeableness, Conscientiousness, Extraversion, Neuroticism, Openness to Experience), relying on data collected among western, educated, industrialized, rich and democratic (WEIRD) populations.

In recent work, Laajaj et al. (2019) suggested that efforts to collect data on these traits in non-WEIRD populations have not travelled well. Using data from 14 surveys from low- and middle-income countries, they document the pattern of lack of reliability (Cronbach-alpha coefficient) and validity (Tucker's congruence coefficient) of the Big Five measures in survey data, concluding that these data fail to capture the intended Big Five structure.[3] In addition to differences in respondents' education (that may diminish the capacity to understand the questionnaire), the authors suggest that this result may reflect differences in the way that such questionnaires are administered: moving from the impersonal data collection instruments typically used in developed countries to questionnaires administered through face-to-face surveys may have increased the importance of acquiescence bias (the tendency to systematically agree or disagree with questions regardless of their content, i.e., always yes/nay saying Baumgartner and Steenkamp (2001)) and of socially desirable response (the tendency for respondents to over-report positive behaviors or under-report negative ones (Edwards, 1953)), while creating room for enumerators to influence respondents' answers in other ways (due to differences in capacity to explain questions, for example). In this article, we use data on the Big Five, collected among a rural population in a developing country through face-to-face interviews, to quantitatively examine the relative importance of these three suggested explanations for the low reliability and validity of the data collected in developing countries.

The data is presented in Section 2. Data on personality traits was collected using the Mini-IPIP questionnaire (Donnellan et al., 2006), a 20-item short form of the 50-item International Personality Item Pool-Five-Factor-Model measure (Goldberg, 1999). Contrary to other short-scale inventories commonly used (including the BFI 15-item scale, on which Laajaj et al (2019) largely rely), the personality inventory we used includes reversed and nonreverse scored questions for all five traits, allowing us to separately correct for acquiescence bias for each trait. Given that, as we will show in this

[2] See also Heckman and Mosso (2014), and Ozer and Benet-Martinez (2006) for reviews.

[3] Following Tavakol and Dennick (2011) we define reliability as the extent to which all the items in a questionnaire measure the same concept and validity as the extent to which a questionnaire measures what it is intended to measure.

section, acquiescence bias is heterogeneous across traits in both size and direction, this apparently minor difference will matter for the adequacy of this correction.

In addition to measures of Big Five personality traits, we also collected data on the importance of social desirability response (SDR). Following Paulhus (2002), we distinguish two dimensions of this behavior: impression management (IM, responses are biased for the purpose of please interviewers) and self-deceptive enhancement (SDE, responses are honest but overly positive). Meyer and Santos (2020) present evidence of the importance of accounting for such multidimensionality in a related context. Finally, we know the identity of all enumerators, which allows us to separately estimate their influence (if any) on the pattern of answers. We discuss, in section 3, how to correct the Big Five for each of these three potential sources of bias.

Section 4 presents the results of our analysis. We quantify the impact of these different potential sources of bias on the psychometric properties of our data, by comparing the corrected measures of personality traits with different benchmarks (including the data analysed by Laajaj et al. (2019) and the data used by Donnellan et al. (2006) to construct the Mini-IPIP), using two measures of reliability and validity: Cronbach's alpha (Cronbach, 1951), a commonly used measure of reliability of a questionnaire, and Tucker's congruence coefficient (Tucker, 1951), a measure of its validity. We show that the bias introduced by yes/nay behavior can be greatly reduced through the use of inventories that allow for the correction of trait-specific acquiescence bias, leading to substantial improvements in the reliability and validity of survey data. However, correcting for SDR or enumerator effects does not significantly improve the psychometric properties of the data once we correct for acquiescence bias.

Section 5 then presents the results of a separate but related analysis: Are the biases hypothesized as drivers of low reliability and validity of the Big Five also contributing to biased estimates of their impact on outcomes of interest? We address this question by estimating the capacity of the Big Five to predict willingness to contribute for a public good, given that the two have often been recognized as being associated with each other (Ozer and Benet-Martinez, 2006). We show that correcting for acquiescence bias doesn't change the Big Five predictive power (a result that follows from the balanced nature of the questionnaire used) but that controlling for SDR and enumerator effects does.

Section 6 presents our conclusions. The use of a balanced scale questionnaire, such as Mini-IPIP, is central for adequate correction of acquiescence bias, and allows us to reach reliable estimates of the structure of personality traits. Other corrections seem less important. When the primary research interest is related with using personality traits as predictors of behavior, the use of a balanced questionnaire inherently addresses concerns about the effect of acquiescence bias on estimates of that effect. However, controlling for social desirability bias and enumerator effects may matter.

2. Data

We use data from a survey conducted, in May 2019, among 560 rural households, randomly selected from village lists in 52 villages in four districts in northern Lao PDR. The data was collected in the context of an evaluation of the impact of a pilot intervention aimed at reducing post-harvest losses due to environmental pests (rodents). Meyer, Santos and Kousonsavath (2020) present further details about this intervention and its outcomes.

In addition to data of demographic variables (including age, gender, formal education and district of residence), we collected data on respondents' personality traits and social desirability response. It was expected that data on the Big Five would help explain some of the behavior during the pilot intervention as well as their willingness to pay for its continuation post-pilot, a question that we will analyse in more detail in Section 5. As with other surveys, respondents were encouraged to answer the surveys in an honest manner, with enumerators emphasizing that there was no right or wrong answer, and that their answers just reflected their personal choices, which would be kept confidential.

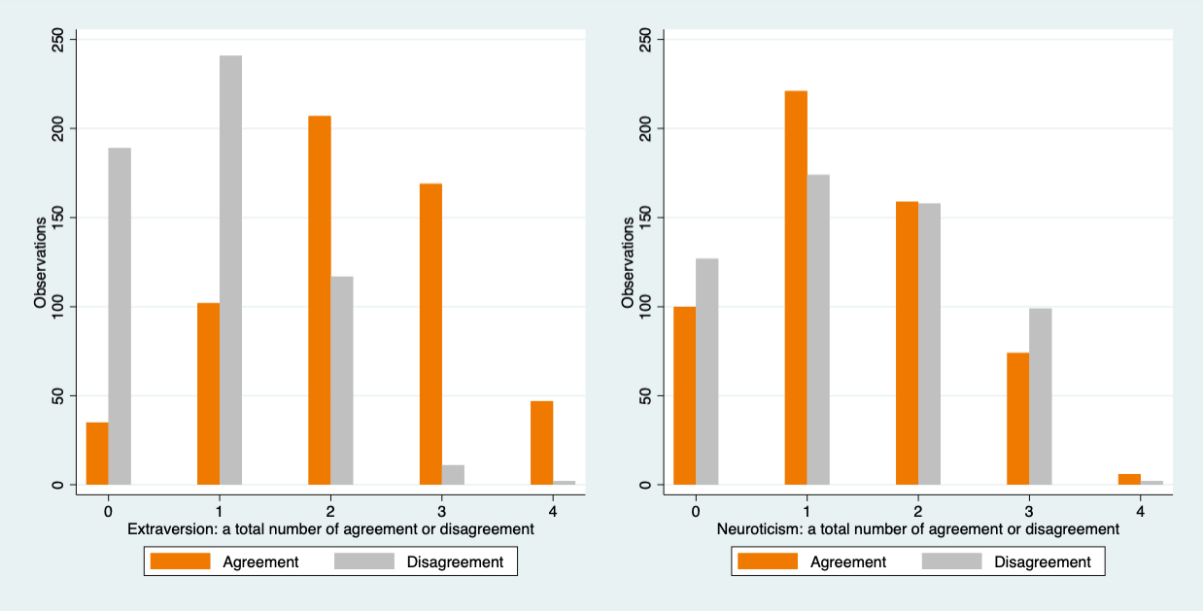
Data on the Big Five were collected using the Mini-IPIP inventory, developed and validated by Donnellan et al. (2006), and consisting of 20 of the 50 standard IPIP-FFM items (Goldberg, 1999).[4] It assesses the Big Five using four items per trait, with two of them scored in the reverse and two in the nonreverse direction (with the exception of openness to experience, where three items are reverse scored). Respondents were asked to state whether they disagreed, neither agree nor disagree or agreed that the statement represented them (with answers scored from one to three, respectively). After recoding the value of reverse scored items, we take the mean over the four questions to construct measures of each personality trait. A higher mean value indicates a higher level of the corresponding trait.

As in other surveys, acquiescence bias, which has been recognized as important factor of variation in the structure of psychometric data (McCrae, Herbst, & Costa Jr, 2001), is present in our data. Using, as an example, the measurement of extraversion, captured using the two items "I am the life of the party" and "I talk to a lot of different people at parties" and the two (reverse scored items) "I don't talk a lot" and "I keep in the background", we would expect that agreement/disagreement with the first two items would elicit disagreement/agreement with the other two. However, as shown in figure 1 (left panel), that is not what we see in the data. Approximately 200 participants agreed or disagreed with more than two of the items designed to measure extraversion, a pattern that suggests the presence of acquiescence bias. Similar patterns are found when measuring other traits, although the importance of acquiescence bias is clearly heterogeneous, as also shown in figure 1: while the importance of agreement or

[4] See Appendix Table A1 for the questions included in this personality scale.

disagreement above the value of two is similar when measuring neuroticism (right panel), respondents exhibit a much greater tendency to agree with statements designed to measure extraversion.[5]

Figure 1: Acquiescence bias is trait-specific: contrasting extraversion and neuroticism



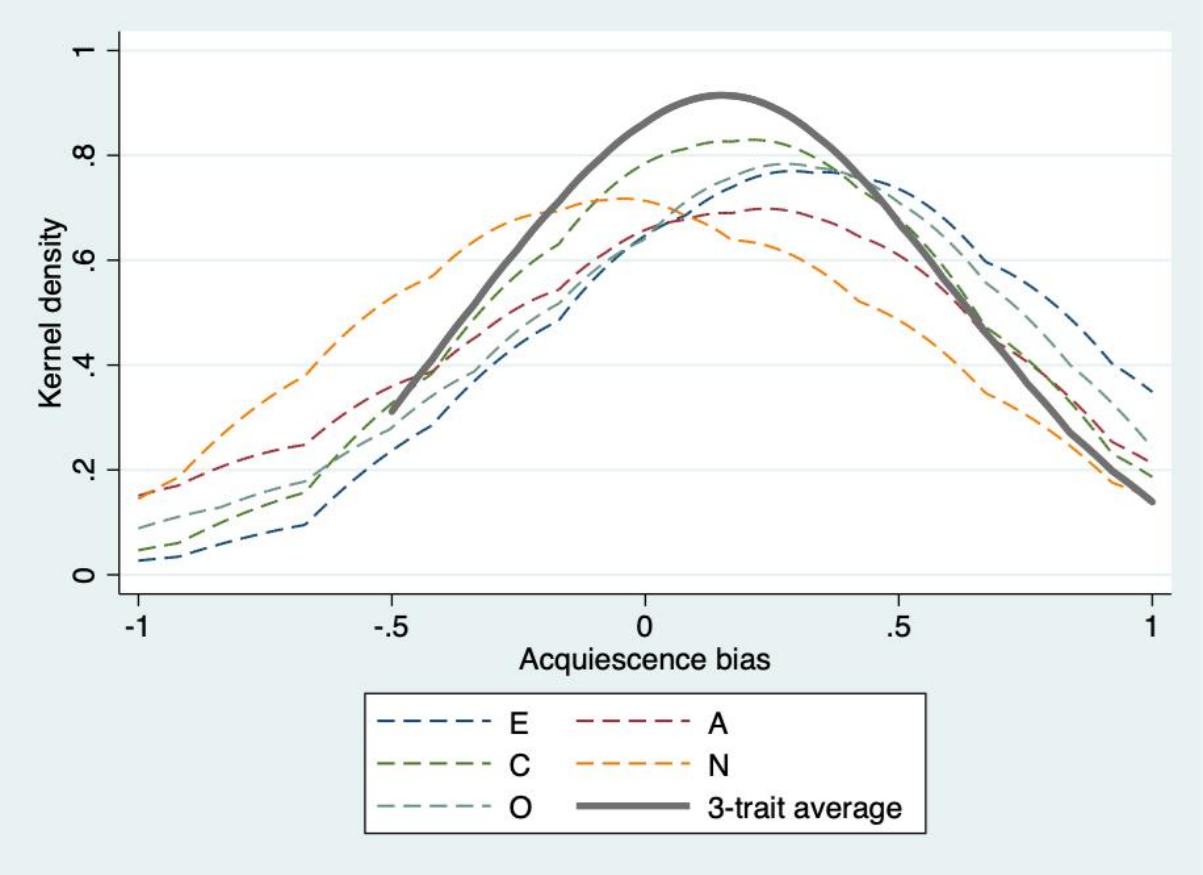
Notes: The figure shows the histograms of response patterns for participants, by personality trait. The x-axis displays the total number of agreement or disagreement, and the y-axis displays the number of observations in the category. Data are from 560 rural households, randomly selected from village lists in 52 villages in four districts in northern Lao PDR. Items designed to measure Extraversion are: "I am the life of the party", "I talk to a lot of different people at parties", "I don't talk a lot (reverse scored)" and "I keep in the background (reverse scored)". Neuroticism is measured using the items "I have frequent mood swings", "I get upset easily", "I am relaxed most of the time (reverse scored)", and "I seldom feel blue (reverse scored)". Respondents were asked to state whether they disagreed, neither agree nor disagree or agreed that the statement describes them. The expectation is that, in the absence of acquiescence bias, agreement/disagreement with the first two items would elicit disagreement/agreement with the two reverse scored items. See Appendix Figure A1 for histograms across five traits.

The inclusion of reverse scored items for all traits is a major advantage of the personality inventory we use as it allows us to calculate the importance of acquiescence bias for each trait separately, avoiding the implicit assumption that such bias is homogenous (in size and direction) across traits (see Table A3 for details). This is not possible when using short-scale inventories such as the BFI-15, used in the surveys analyzed by Laajaj et al. (2019): in that inventory, the absence of reverse scored items for two traits (agreeableness and openness) forces the authors to use the average acquiescence bias across the other three traits (conscientiousness, extraversion, and emotional stability) to correct for its presence in

[5] See also Appendix Figure A1 and Table A2 for a quantification of acquiescence bias across the five traits.

all five traits. Figure 2, which presents the kernel density estimates of trait-specific acquiescence bias and of its average across conscientiousness, extraversion, and emotional stability (i.e., the three traits for which acquiescence bias can be properly corrected when using the BFI 15), illustrates both the heterogeneity of such bias and the inadequacy of relying on averages across traits for this correction: for example, while the estimated 3-trait average acquiescence bias ranges from -0.5 to 1, the lower bound for trait-specific acquiescence bias is always -1. [6]

Figure 2: Kernel density estimates of acquiescence bias.



Notes: The figure shows the Kernel density estimates of acquiescence bias, by personality trait. The x-axis displays the value of the estimated acquiescence bias (see Section 3 for details), and the y-axis displays the estimated density. Data are from 560 rural households, randomly selected from village lists in 52 villages in four districts in northern Lao PDR. See Appendix Table A3 for descriptive statistics (mean value and standard deviation) of acquiescence bias across the five traits.

In addition to data on the Big Five, we collected data on SDR using the 16-item BIDR (Balanced Inventory of Desirable Responding) developed by Hart, Ritchie, Hepper, and Gebauer (2015), which takes into account the potential multidimensionality of the tendency to provide socially desirable

[6] See Appendix Table A2 for the detailed distribution of overall response patterns.

responses.[7] In addition to controlling for Impression Management (the responses are biased for the purpose to please others), the dimension of SDR captured in commonly used measures such as the Marlow–Crowne Scale, the BIDR allows us to control for Self-Deceptive Enhancement (the responses are honest, but they are overly positive). The BIDR includes eight items designed to capture each of these dimensions, with four in each group being reverse scored. Respondents are asked to state their agreement with the statement using a Likert Scale, where one refers to strongly disagree, and five represents strongly agree. Once we recode the value of all reverse scored items, we estimate each of the dimensions of SDR bias as the mean value of the respective eight questions. A higher mean value indicates a higher degree of the tendency to respond in a socially desirable way. Finally, we know the identity of all enumerators, which allows us to separately identify for their influence on the pattern of answers.

We will compare our data with three benchmarks. The first two are the survey and online data analyzed in Laajaj et al. (2019, p3), mostly collected among urban populations in 14 developing countries (including Lao PDR). In addition, we contrast our data with the one used by Donnellan et al. (2006) to develop and validate the Mini-IPIP inventory. The availability of this last dataset allows us to evaluate the similarity between data structures of the same questionnaire applied to different populations using Tucker's congruence coefficient (Tucker, 1951).

Descriptive statistics for the different datasets are presented in Table 1. In addition to the differences in administration methods, we note that respondents in our sample are more likely to live in rural regions and to be older and less educated than the set of populations analyzed by Laajaj et al. (2019), including in Lao PDR. These differences may contribute to explain differences in reliability and validity of the data, hence we include them as additional covariates in our analysis.

[7] See Appendix Table A4 for the questions included in this questionnaire.

Table 1: Descriptive Statistics of Datasets

Dataset	Our data	STEP Surveys – Lao PDR (Laajaj et al. 2019)	STEP Surveys (Laajaj et al. 2019)	Online Surveys (Laajaj et al. 2019)	MINI-IPIP Validated Data (Donnellan et al., 2006)
Country	Lao PDR	Lao PDR	14 developing countries	14 developing countries	USA
Rural Regions (%)	100%	66%	11%	N/A	0%
N	560	2,793	40,584	198,356	2,663
Big Five Inventory	MINI-IPIP 20	BFI 15	BFI 15	BFI 15	MINI-IPIP 20
Number of Items	20	15	15	15	20
Balanced Scale across Five Traits	Yes	No	No	No	Yes
Administration	Face-to-Face	Face-to-Face	Face-to-Face	Online	Face-to-Face
Self-administered	No	No	No	Yes	Yes
Age	42	36	37	23	19[a]
Gender (% Female)	32%	54%	59%	70%	64%
Education	4.8 Years	6.1 Years	10.8 Years	81% Respondents with College Education	Undergraduate Students

Notes: [a] the average age in MINI-IPIP Validated Data is estimated based on the authors' note that "all respondents are the first-year college students, and 97% of the sample was either 18 or 19 years of age."

3. Accounting for respondents' biases in face-to-face surveys

We consider the potential importance of three sources of bias in the measurement of personality traits: acquiescence bias, social desirability response and enumerator effects. As it is standard in the psychology literature (Rammstedt, Kemper, & Borg, 2013; Soto, John, Gosling, & Potter, 2008), we use acquiescence bias correction as the foundation for other bias corrections.

Our approach to correct for acquiescence bias is similar to Laajaj et al. (2019) but benefits from a richer questionnaire where each personality trait is measured by both reverse and nonreverse coded items (two each, except for the trait of Openness, where three items are reverse scored). After first recoding the answers to reverse scored items (RSI), we estimate the trait-specific acquiescence bias (AB) at the individual *I* level as follows:

$$\mathrm{AB_{Pi}} = \frac{\overline{\mathrm{NRSI_{Pi}}} - \overline{\mathrm{RSI_{Pi}}}}{2} \quad , \quad \mathrm{P} = \mathrm{A, C, E, N, O} \tag{1}$$

where P stands for each of the Big Five personality traits of individual *i* (Agreeableness, Conscientiousness, Extraversion, Neuroticism, Openness to experience, with initial capitalized to match Eq. (1)), and $\overline{\mathrm{RSI}}$ and $\overline{\mathrm{NRSI}}$ correspond to the average score for reverse or nonreverse scored items for each trait, respectively. As shown in equation (1), each trait's AB equals the average value of the difference between the mean value of NRSI and that of RSI. We correct for this bias, at trait-individual level, by adding our calculation of AB from equation (1) to every reverse scored item and subtracting that same value from every nonreverse scored item, as shown below:

$$\mathrm{NRSI_{Pni}^{AB}} = \mathrm{NRSI_{Pni}} - \mathrm{AB_{Pi}} \tag{2.1}$$

$$\mathrm{RSI_{Pni}^{AB}} = \mathrm{RSI_{Pni}} + \mathrm{AB_{Pi}} \tag{2.2}$$

The intuition behind this approach is simple. In the absence of acquiescence bias, a respondent's answer to nonreverse and (recoded) reverse scored items should be equivalent. For instance, in the absence of acquiescence bias, a respondent who agreed (score of 3) with the argument "I don't talk a lot" would disagree (score of 1) with the reverse argument "I talk to a lot with different people at parties" (and vice versa). After recoding the reverse scored item, this respondent should have a score of 3 in both items. If s/he does not, the acquiescence bias correction adjusts each item's score in order to make them equivalent. As shown in the figures 1 and 2 above (see also Table A3), our estimates of trait-specific acquiescence bias do not support the assumption that such biases are homogeneous.

In the next step, we control for social desirability bias and enumerator effects at the level of individual *i*. We use ordinary least-squares regression to estimate the relation between each of 20 items of the Mini-IPIP ($\mathrm{I_n}$, n = 1, 2, 3 … 20), now corrected for acquiescence bias, on measures of Impression Management (IM) and Self-Deceptive Enhancement (SDE) and a full set of enumerator fixed effects

(E_j), while controlling for respondents' demographic characteristics (age, gender, formal education, and district fixed effects (X_i)).

$$I_{ni} = \alpha + \gamma IM_i + \eta SDE_i + \psi_j E_{(ji)} + X_i'\delta + \epsilon_i , \qquad n = 1, 2, 3 \dots 20 \tag{3}$$

With these estimates, we can adjust for the importance of each of these biases by subtracting the associated average bias from the response to each item I_n.

$$I_{ni}^{Ej} = I_{ni} - \psi_j \tag{3.1}$$

$$I_{ni}^{SDR} = I_{ni} - (\gamma * IM_i) - (\eta * SDE_i) \tag{3.2}$$

$$I_{ni}^{SDR\ \&\ Ej} = I_{ni} - (\gamma * IM_i) - (\eta * SDE_i) - \psi_j \tag{3.3}$$

4. Validity and reliability of personality structure

We assess the importance of the corrections discussed in the previous section in terms of reliability and validity of the data on personality traits using two indicators. The first is Cronbach's alpha (Cronbach, 1951; Tavakol & Dennick, 2011), which evaluates the reliability of questionnaires and ranges from 0 to 1, with a higher value indicating that the multi-item indicator measures a common concept. The second in Tucker's congruence coefficient (Lorenzo-Seva & Ten Berge, 2006; Tucker, 1951), a validity indicator used to measure the similarity between data structure (factor loadings derived from the principal factor analysis) of the same questionnaire applied to two different populations (here, the similarity of the Mini-IPIP inventory applied to the population in our study and to US college students collected by Donnellan et al. (2006)). The question we want to answer is which corrections (if any) matter most in terms of bringing our data closer to the different possible benchmarks.[8]

Table 2 presents our main results in terms of psychometric assessments. Panel A presents the reliability and validity statistics for the raw data. Panel B presents the same statistics when we first correct for acquiescence bias using the approach followed in Laajaj et al., (2019) and, subsequently, correct for socially desirable response, enumerator effects or both. Finally, panel C presents the results when we control for trait-specific acquiescence bias, and then for the importance of socially desirable response and for enumerator effects or both. Panel D presents, for ease of comparison, the same statistics calculated in the benchmark datasets.

[8] In the traditional long-scale inventory, with over 50 items, a Cronbach alpha of 0.7 is often considered a minimum threshold for high consistency (Tavakol & Dennick, 2011) while a congruence coefficient of at least 0.85 corresponds to a fair similarity (Lorenzo-Seva & Ten Berge, 2006). Those thresholds are not valid when using short-scale inventories that rely on a relatively small number of items (Gosling, Rentfrow, & Swann Jr, 2003; Streiner, 2003).

Table 2: Evaluating the importance of biases on the reliability and validity of personality traits

	Cronbach's Alpha	Tucker's congruence coefficient
Panel A: Raw data		
	0.15	0.48
Panel B: Acquiescence bias (average of three traits)		
AB. only	0.27	0.42
AB. and SDR	0.26	0.41
AB. and Enumerator	0.24	0.55
AB., SDR and Enumerator	0.23	0.55
Panel C: Acquiescence bias (trait-specific estimates)		
AB. only	0.51	0.64
AB. and SDR	0.50	0.64
AB. and Enumerator	0.50	0.67
AB, SDR and Enumerator	0.49	0.72
Panel D: Statistics in benchmark datasets		
Lao PDR (STEP Survey) [a]	0.38	0.70
STEP Surveys (all) [a]	0.49	0.73
Online Survey [a]	0.57	0.90
MINI-IPIP Validated Data [b]	0.70	1

Notes: Panel A presents the reliability (Cronbach's alpha) and validity (Tucker's congruence coefficient) statistics for the raw data. Panel B presents the same statistics when we correct for acquiescence bias (AB) using the approach followed in Laajaj et al. (2019), and subsequently correct for socially desirable response (SDR), enumerator fixed effects or both. Panel C presents the results when we correct for trait-specific acquiescence bias and, subsequently, for socially desirable response and enumerator effects or both. Panel D presents the same statistics calculated for all the benchmark datasets.
[a] Estimated by Laajaj et al., (2019, p4). [b] Estimated by Donnellan et al., (2006, p194)

These results support two main conclusions. Firstly, acquiescence bias is the most important source of bias in these measures in our data. As shown in panels B or C, once we control for acquiescence bias, accounting for the effect of the other two potential sources of bias does little to improve the reliability and validity of the data. Secondly, the values of reliability and validity measures are much closer to the benchmark datasets when we can estimate trait-specific acquiescence bias (panel C), than when we use an average across traits (as in panel B).

In short, our analysis shows that properly correcting for acquiescence bias in the Big Five requires a balanced questionnaire, such as the Mini-IPIP 20-item scale which includes both nonreverse and reverse scored items for all traits. Correcting for trait-specific acquiescence bias allows us to reach satisfactorily reliable estimates of the structure of personality data collected in developing countries. However,

correcting for SDR or enumerator effects does not significantly improve the psychometric properties of the data once we correct for acquiescence bias.

5. Analyzing the predictive power of the Big Five

The previous section showed that a balanced questionnaire allows us to overcome Laajaj et al (2019) critique of the feasibility of capturing the structure of personality traits in one non-WEIRD population. However, most economists will be interested in a different (if potentially related) question: which corrections (if any) matter most in terms of shaping our conclusions about the effects of personality traits on some outcome of interest. Given that personality traits are often recognized as being associated with pro-social behavior (Ozer and Benet-Martiez 2006), we address this question by exploring the predictive power of the Big Five, after the different corrections discussed in the previous section, on willingness to contribute to fund a public good.

As part of the evaluation of the impact of the pilot intervention studied in Meyer, Santos and Kousonsavath (2020), respondents were asked about their willingness to contribute to fund the intervention in the future. Table 3 presents five different values of the association between the Big Five and respondents' (log)WTP, estimated using OLS.

In Column 1 we use the raw data (i.e., without correcting for any source of bias). We then progress by sequentially correcting for trait-specific acquiescence bias (column 2), socially desirable response (column 3), enumerator fixed effects (column 4) and both socially desirable response and enumerator effects (column 5). We control for respondents' demographic characteristics (including age, gender and formal education), wealth (including livestock (in TLU) and crop land (in ha)) and for district fixed effects in all specifications. Standard errors are clustered at the village level.

Table 3: WTP for public good: OLS estimates

Dep Var: WTP for public good (Log)	(1)	(2)	(3)	(4)	(5)
Extraversion	0.863**	0.962***	0.715**	0.901**	0.909**
	(0.336)	(0.328)	(0.349)	(0.391)	(0.391)
Agreeableness	-0.293	-0.266	-0.157	0.0967	0.0780
	(0.296)	(0.294)	(0.283)	(0.287)	(0.288)
Conscientiousness	0.954**	0.919**	1.316***	0.606	0.688*
	(0.413)	(0.412)	(0.430)	(0.376)	(0.378)
Neuroticism	1.249***	1.304***	1.124***	0.265	0.189
	(0.419)	(0.423)	(0.400)	(0.393)	(0.388)
Openness	1.037***	0.860***	0.782**	-0.151	-0.102
	(0.323)	(0.307)	(0.326)	(0.298)	(0.297)
Constant	-0.455	-0.427	5.912***	1.532	3.356
	(1.889)	(1.996)	(2.100)	(2.380)	(2.617)
Demographic controls	Yes	Yes	Yes	Yes	Yes
Wealth	Yes	Yes	Yes	Yes	Yes
Trait-specific correction for acquiescence bias	No	Yes	Yes	Yes	Yes
Socially desirable response	No	No	Yes	No	Yes
Enumerator fixed effects	No	No	No	Yes	Yes
H_0: X(1) = X(c) , c = 2-5 [a]	N/A	0.0319	0.0303	2.092*	2.048*
	(N/A)	(1.166)	(1.167)	(1.140)	(1.147)
H_0: ΣBig Five = 0 [b]	3.810****	3.778****	3.780****	1.718**	1.762**
	(0.808)	(0.841)	(0.842)	(0.804)	(0.815)
N	560	560	560	560	560
Adjusted R^2	0.077	0.074	0.104	0.329	0.331

Notes: The dependent variable is the monetary value of WTP for public good. All regressions control for respondents' demographic characteristics (age, gender and formal education), wealth (livestock (in TLU) and crop land (in ha)) and district fixed effects. Column (1) presents the estimates of the association between personality traits when we use the raw data. Column (2) presents estimates when personality data is corrected for trait-specific acquiescence bias. Columns (3) to (5) present the results when we correct for trait-specific acquiescence bias, the importance of socially desirable response, enumerator fixed effects or both. Standard errors in parentheses are clustered at the village level. Full results are presented in Appendix Table A6. * p < 0.1, ** p < 0.05, *** p < 0.01, **** p < 0.001

[a] The null hypothesis is whether the joint predictive power of the Big Five is indistinguishable from the estimates presented in column (1).

[b] The null hypothesis is whether the sum of the effect of the five personality traits is statistically different from zero.

A comparison of the estimates obtained using the raw data (column (1)) with those presented in column (2) supports our first conclusion: correcting for acquiescence bias has little impact on Big Five's predictive power on WTP. This result follows from the balanced structure of the Mini-IPIP questionnaire and the nature of the acquiescence bias correction which, as shown in equation (1), will leave the average value of each trait unchanged at individual level (again, with the exception of the trait Openness to Experience).[9]

In contrast with this result, controlling for social desirability bias (column (3)) does change the relative importance of some of the personality traits, although the estimates are globally similar (in the sense that they are statistically identical to those presented in column (1)). Larger differences emerge when we control for enumerator fixed effects (column (4)): not only are individual estimates different, with only the effect of Extraversion precisely estimated at traditional levels of statistical significance, we are now able to reject the null hypothesis that such correction makes no difference in terms of the joint effect of the Big Five. Finally, when we control for all three sources of bias, the results largely mimic the estimates presented in column (4), although Conscientiousness is now a significant predictor of WTP (although only at the 10% level) and the magnitude of the effect of each trait is also significantly smaller. Overall, we can conclude that controlling for both social desirability bias and, especially, for enumerator effects matters when estimating the predictive power of the Big Five traits, while the acquiescence bias correction has a limited impact when using an appropriately chosen questionnaire.

6. Conclusion

This article addresses Laajaj et al (2019) critique of the reliability and validity (and, implicitly, predictive power) of data on Big Five personality trait collected in non-WEIRD societies by investigating the relative importance of three sources of bias: acquiescence bias, social desirability bias, and enumerator interactions. Using data from face-to-face asked questionnaires, our analysis shows that the inclusion of reverse and non-reverse items for all traits (i.e., the use of a balanced questionnaire such as the Mini-IPIP, but not the BFI-1) allows for an acceptable correction of trait-specific acquiescence bias and allows us to overcome the critique that motivated this analysis. Moreover, when the primary research interest is obtaining estimates of the importance of the Big Five as predictors of life outcomes, using a balanced questionnaire inherently addresses concerns about the importance of acquiescence bias on those estimates. However, social desirability bias and enumerator effects may still be important.

[9] Comparisons of Big Five's descriptive statistics (mean and standard error) after different bias correction can be found in Appendix Table A5.

References

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of marketing research, 38*(2), 143-156.

Burks, S. V., Lewis, C., Kivi, P. A., Wiener, A., Anderson, J. E., Götte, L., . . . Rustichini, A. (2015). Cognitive skills, personality, and economic preferences in collegiate success. *Journal of Economic Behavior & Organization, 115*, 30-44.

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

Donato, K., Miller, G., Mohanan, M., Truskinovsky, Y., & Vera-Hernández, M. (2017). Personality traits and performance contracts: Evidence from a field experiment among maternity care providers in India. *American Economic Review, 107*(5), 506-510.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192.

Drouvelis, M., & Georgantzis, N. (2019). Does revealing personality data affect prosocial behaviour? *Journal of Economic Behavior & Organization, 159*, 409-420.

Dupuy, A., & Galichon, A. (2014). Personality traits and the marriage market. *Journal of Political Economy, 122*(6), 1271-1319.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of applied Psychology, 37*(2), 90.

Gensowski, M. (2018). Personality, IQ, and lifetime earnings. *Labour Economics, 51*, 170-183.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe, 7*(1), 7-28.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528.

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open, 5*(4), 2158244015621113.

Laajaj, R., Macours, K., Hernandez, D. A. P., Arias, O., Gosling, S. D., Potter, J., . . . Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science advances, 5*(7), eaaw5226.

Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57-64.

McCrae, R. R., Herbst, J. H., & Costa Jr, P. T. (2001). Effects of acquiescence on personality factor structures. *Personality and temperament: Genetics, evolution, and structure*, 217-231.

Meyer, S., & Santos, P. (2020). Correcting for bias using multidimensional measures of Socially Desirable Response. *Economics Letters, 196*, 109574.

Meyer, S., Santos, P. & Kousonsavath (2020): Using contests to promote coordinated control of invasive species: an experimental evaluation. *Working Paper*.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol., 57*, 401-421.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. *The role of constructs in psychological and educational measurement, 49459*.

Proto, E., Rustichini, A., & Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy, 127*(3), 1351-1390.

Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting big five personality measurements for acquiescence: An 18–country cross–cultural study. *European Journal of Personality, 27*(1), 71-81.

Savelyev, P. A., & Tan, K. T. (2019). Socioemotional skills, education, and health-related outcomes of high-ability individuals. *American Journal of Health Economics, 5*(2), 250-280.

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.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99-103.
Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, *2*, 53.
Tucker, L. R. (1951). *A method for synthesis of factor analysis studies*. Educational Testing Service Princeton NJ.

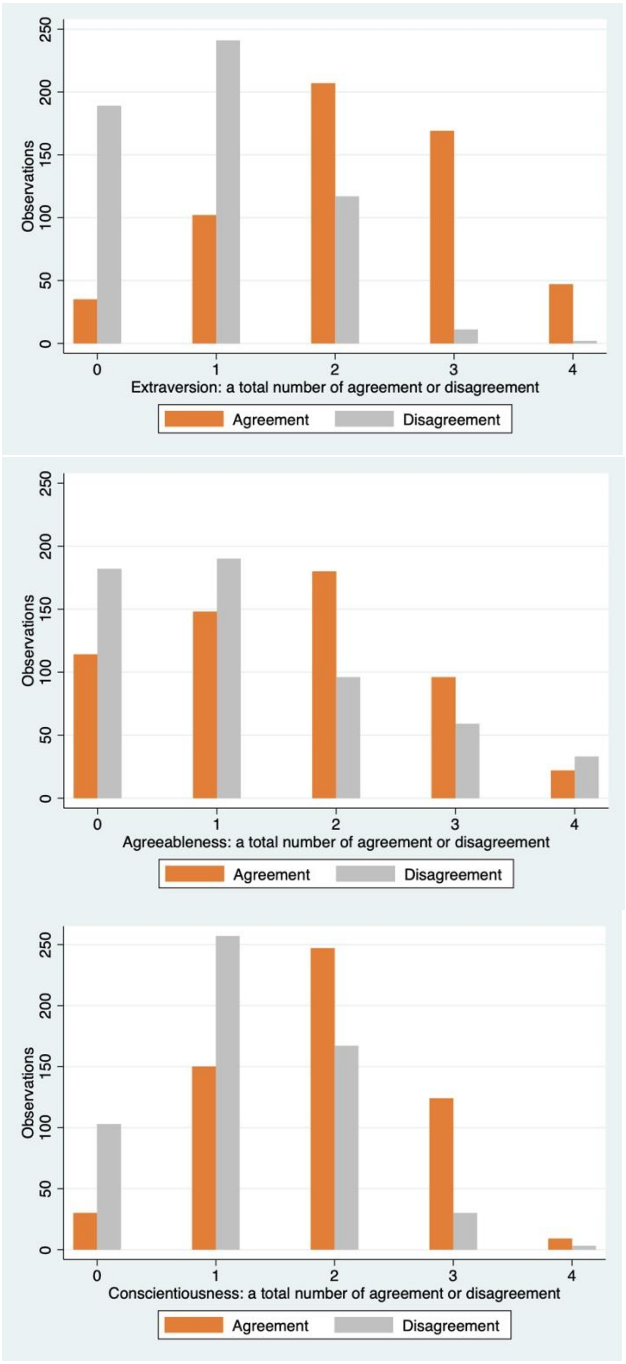
Appendix for

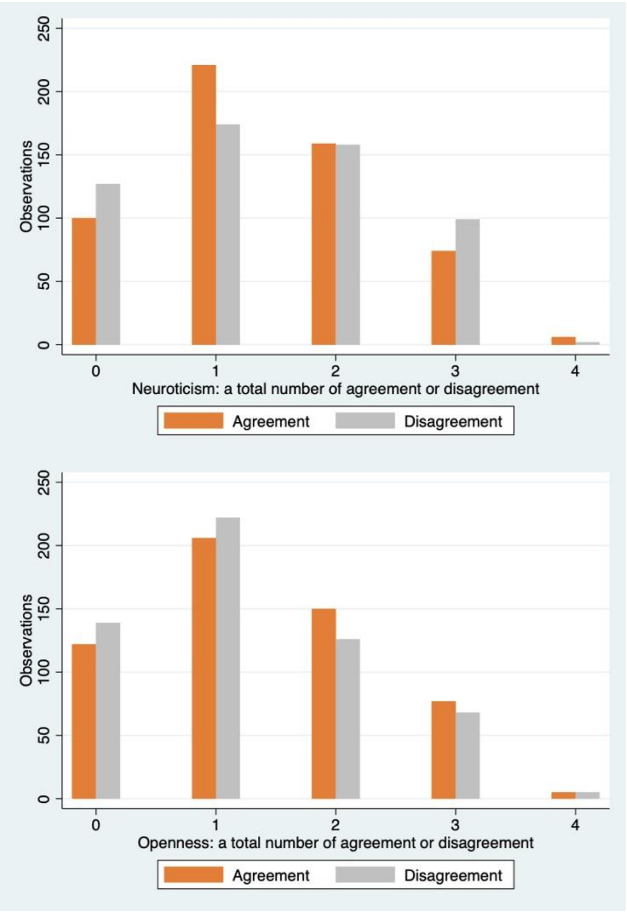
Improving the reliability and validity of data on personality traits in developing countries

An Huang*, Paulo Santos

*Corresponding author. Email: an.huang@monash.edu

Figure A1. Acquiescence bias is trait-specific: contrasting five traits





Notes: The figure shows the histograms of response patterns for participants grouped by the personality trait. The x-axis displays a total number of agreement or disagreement, and the y-axis displays the number of observations in the category. Data are from 560 rural households, randomly selected from village lists in 52 villages in four districts in northern Lao PDR. See Appendix Figure A1 for detailed statements across five traits. Respondents were asked to state whether they disagreed, neither agree nor disagree or agreed that the statement, and we would expect that agreement/disagreement with the first two items would elicit disagreement/agreement with the other two reverse scored items.

Table A1. MINI-IPIP 20 Questionnaire.

Item	Item	Trait	Keyed
1	I am the life of the party.	E	
2	I sympathize with others' feelings.	A	
3	I get chores done right away.	C	
4	I have frequent mood swings.	N	
5	I have a vivid imagination.	I	
6	I don't talk a lot	E	R
7	I am not interested in other people's problems.	A	R
8	I often forget to put things back in their proper place.	C	R
9	I am relaxed most of the time.	N	R
10	I am not interested in abstract ideas.	I	R
11	I talk to a lot of different people at parties.	E	
12	I feel others' emotions	A	
13	I like order.	C	
14	I get upset easily.	N	
15	1 have difficulty understanding abstract ideas.	I	R
16	I keep in the background.	E	R
17	I am not really interested in others.	A	R
18	I make a mess of things.	C	R
19	I seldom feel blue.	N	R
20	I do not have a good imagination.	I	R

Notes: E, A, C, N and O are abbreviations for Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience, respectively. R stands for reverse scored Items.

Table A2. Distribution of Overall Response Pattern in the Big Five Questionnaire.

Number of responses:	Agree	Neither Agree nor Disagree	Disagree
0	0	30	6
1	3	34	16
2	6	66	48
3	19	61	77
4	32	51	63
5	58	62	60
6	54	42	69
7	63	46	52
8	64	45	51
9	65	30	31
10	46	35	19
11	48	24	18
12	44	15	22
13	21	13	13
14	19	4	13
15	11	1	2
16	2	0	0
17	5	1	0
N	560	560	560

Notes: This table represents the distribution of the overall response pattern. Given that the inventory includes, in total, nine nonreverse and eleven reverse scored items, the fact that several respondents agreed or disagreed with more than eleven items is an indication of the presence of acquiescence bias in this data. Using this rough cut-off, 102/560 respondents agreed with more statements than what would be consistent with the structure of the inventory.

Table A3. Descriptive Statistics of Trait-specific Acquiescence Bias.

	Acquiescence bias		
Trait	Mean	Standard Deviation	N
Extraversion	0.310	0.397	560
Agreeableness	0.086	0.495	560
Conscientiousness	0.160	0.357	560
Neuroticism	-0.004	0.428	560
Openness to Experience	0.184	0.429	560

Notes: This table displays the mean value and standard deviation of acquiescence bias across five traits. Data are from 560 rural households, randomly selected from village lists in 52 villages in four districts in northern Lao PDR. See Section 3 for estimation details.

Table A4. Social Desirability Response Questionnaire.

Item		Scale	Scored
1	I have not always been honest with myself.	SDE	R
2	I always know why I like things.	SDE	
3	It's hard for me to shut off a disturbing thought.	SDE	R
4	I never regret my decisions.	SDE	
5	I sometimes lose out on things because I can't make up my mind soon enough.	SDE	R
6	I am a completely rational person.	SDE	
7	I am very confident of my judgments	SDE	
8	I have sometimes doubted my ability as a lover.	SDE	R
9	I sometimes tell lies if I have to.	IM	R
10	I never cover up my mistakes.	IM	
11	There have been occasions when I have taken advantage of someone	IM	R
12	I sometimes try to get even rather than forgive and forget.	IM	R
13	I have said something bad about a friend behind his/her back	IM	R
14	When I hear people talking privately, I avoid listening.	IM	
15	I never take things that don't belong to me.	IM	
16	I don't gossip about other people's business.	IM	

Notes: SDE and IM stand for Self-Deceptive Enhancement and Impression Management, respectively.
R stands for reverse scored Items.

Table A5. Descriptive Statistics of the Big Five.

	Raw	Trait-specific acquiescence bias		AB. and Social Desirability Bias		AB. and Enumerator effects		AB., SDR., and Enumerator effects	
Big Five	*Mean [SD]*	*Mean [SD]*	*t (p-value)*	*Mean [SD]*	*t (p-value)*	*Mean [SD]*	*t (p-value)*	*Mean [SD]*	*t (p-value)*
Extraversion	2.341 [0.392]	2.341 [0.392]	n/a (n/a)	2.343 [0.388]	-2.751 (0.0061)	2.244 [0.353]	14.73 (<0.001)	2.246 [0.352]	13.44 (<0.001)
Agreeableness	1.928 [0.448]	1.928 [0.448]	n/a (n/a)	1.899 [0.450]	40.24 (<0.001)	1.970 [0.392]	-4.373 (<0.001)	1.940 [0.391]	-1.310 (0.191)
Conscientiousness	2.308 [0.365]	2.308 [0.365]	n/a (n/a)	1.894 [0.357]	194.1 (<0.001)	2.121 [0.322]	26.68 (<0.001)	1.706 [0.318]	78.74 (<0.001)
Neuroticism	1.809 [0.398]	1.809 [0.398]	n/a (n/a)	2.484 [0.394]	-2.6e+02 (<0.001)	1.810 [0.348]	-0.0945 (0.925)	2.485 [0.342]	-78.49 (<0.001)
Openness to Experience	2.223 [0.400]	2.315 [0.429]	-10.138 (<0.001)	2.254 [0.418]	-3.331 (<0.001)	2.360 [0.366]	-12.60 (<0.001)	2.300 [0.362]	-6.519 (<0.001)
N	560	560		560		560		560	

Notes: The t-test refers to comparisons of means between raw and bias corrected Big Five personality traits.

Table A6: OLS estimates of WTP for public good

Dep Var: WTP for public good (Log)	(1)	(2)	(3)	(4)	(5)
Extraversion	0.863**	0.962***	0.715**	0.901**	0.909**
	(0.336)	(0.328)	(0.349)	(0.391)	(0.391)
Agreeableness	-0.293	-0.266	-0.157	0.0967	0.0780
	(0.296)	(0.294)	(0.283)	(0.287)	(0.288)
Conscientiousness	0.954**	0.919**	1.316***	0.606	0.688*
	(0.413)	(0.412)	(0.430)	(0.376)	(0.378)
Neuroticism	1.249***	1.304***	1.124***	0.265	0.189
	(0.419)	(0.423)	(0.400)	(0.393)	(0.388)
Openness	1.037***	0.860***	0.782**	-0.151	-0.102
	(0.323)	(0.307)	(0.326)	(0.298)	(0.297)
Age	0.0209*	0.0207*	0.0249**	0.0111	0.0124
	(0.0107)	(0.0107)	(0.0101)	(0.00963)	(0.00975)
Gender	-0.637**	-0.631**	-0.588**	-0.400	-0.380
	(0.302)	(0.304)	(0.298)	(0.282)	(0.282)
Schooling	0.0894*	0.0877*	0.0948*	0.0970*	0.0996*
	(0.0533)	(0.0521)	(0.0519)	(0.0525)	(0.0532)
TLU	-0.00107	-0.00241	0.00101	0.00337	0.00372
	(0.0256)	(0.0258)	(0.0235)	(0.0246)	(0.0241)
Crop Land (ha)	0.0640	0.0656	0.0799	0.0873*	0.0904*
	(0.0546)	(0.0546)	(0.0514)	(0.0473)	(0.0462)
IM			-0.878***		0.0891
			(0.293)		(0.369)
SDE			-1.015**		-0.718*
			(0.409)		(0.381)
Constant	-0.455	-0.427	5.912***	1.532	3.356
	(1.889)	(1.996)	(2.100)	(2.380)	(2.617)
Demographic controls	Yes	Yes	Yes	Yes	Yes
Wealth	Yes	Yes	Yes	Yes	Yes
Trait-specific correction for acquiescence bias	No	Yes	Yes	Yes	Yes
Socially desirable response	No	No	Yes	No	Yes
Enumerator fixed effects	No	No	No	Yes	Yes

H_0: X(1) = X(c) , c = 2-5 [a]	N/A	0.0319	0.0303	2.092*	2.048*
	(N/A)	(1.166)	(1.167)	(1.140)	(1.147)
H_0: ΣBig Five = 0 [b]	3.810****	3.778****	3.780****	1.718**	1.762**
	(0.808)	(0.841)	(0.842)	(0.804)	(0.815)
N	560	560	560	560	560
Adjusted R^2	0.077	0.074	0.104	0.329	0.331

Notes: For details on the regressions, see the notes for Table 3.