



Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS



Christopher J. Soto^{a,*}, Oliver P. John^b

^a Department of Psychology, Colby College, United States

^b Department of Psychology, University of California, Berkeley, United States

ARTICLE INFO

Article history:

Received 12 September 2016

Revised 13 January 2017

Accepted 22 February 2017

Available online 24 February 2017

Keywords:

Big Five

Five-factor model

Facets

Personality measurement

Short measures

ABSTRACT

The Big Five Inventory–2 (BFI-2) uses 60 items to hierarchically assess the Big Five personality domains and 15 more-specific facet traits. The present research develops two abbreviated forms of the BFI-2—the 30-item BFI-2-S and the 15-item BFI-2-XS—and then examines their measurement properties. At the level of the Big Five domains, we find that the BFI-2-S and BFI-2-XS retain much of the full measure's reliability and validity. At the facet level, we find that the BFI-2-S may be useful for examining facet traits in reasonably large samples, whereas the BFI-2-XS should not be used to assess facets. Finally, we discuss some key tradeoffs to consider when deciding whether to administer an abbreviated form instead of the full BFI-2.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

1.1. The BFI-2-S and BFI-2-XS: Short and extra-short forms of the Big Five Inventory–2

Individual differences in people's characteristic patterns of thinking, feeling, and behaving can be organized in terms of the Big Five personality trait domains (Goldberg, 1993; John, Naumann, & Soto, 2008; McCrae & Costa, 2008). Moreover, these five broad domains can be conceptualized hierarchically, with each domain subsuming several more-specific facet traits (DeYoung, Quilty, & Peterson, 2007; McCrae & Costa, 2010; Roberts, Chernyshenko, Stark, & Goldberg, 2005). The Big Five Inventory–2 (BFI-2; Soto & John, in press) is a 60-item questionnaire that operationalizes this hierarchical conceptualization of personality structure by assessing the Big Five domains and 15 facets: Extraversion (with facets of Sociability, Assertiveness, and Energy Level), Agreeableness (Compassion, Respectfulness, and Trust), Conscientiousness (Organization, Productiveness, and Responsibility), Negative Emotionality (Anxiety, Depression, and Emotional Volatility), and Open-Mindedness (Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination). The present research was conducted to (a) develop a 30-item short form (the BFI-2-S) and a 15-item

extra-short form (the BFI-2-XS) of the BFI-2, (b) examine the extent to which these short forms retain the reliability and validity of the full BFI-2, and (c) test whether the BFI-2 short forms should only be used to assess personality at the level of the Big Five domains, or whether they are also appropriate for examining facet-level traits.

1.2. The Big Five Inventory–2 and the need for short forms

The BFI-2 has some important psychometric strengths. First, it has a conceptually coherent and empirically robust hierarchical structure, with three facets nested within each Big Five domain (Soto & John, in press). This hierarchical measurement model helps address the bandwidth-fidelity tradeoff: the phenomenon that broadly defined traits tend to predict a wider range of criteria, whereas narrowly defined traits tend to predict closely aligned criteria more accurately (Cronbach & Gleser, 1957; John, Hampson, & Goldberg, 1991). By balancing descriptive breadth at the domain level with specificity at the facet level, the BFI-2's hierarchical structure enhances its power to accurately predict a wide range of external criteria (Soto & John, in press).

Second, the BFI-2 effectively minimizes the influence of acquiescent responding: the tendency of some individuals to consistently agree (yea-saying) or disagree (nay-saying) with items regardless of their content (Jackson & Messick, 1958). Uncontrolled individual differences in acquiescence can bias the results of analyses conducted at both the scale and item levels; for example, they can distort a measure's factor structure (Rammstedt & Farmer,

* Corresponding author at: Department of Psychology, Colby College, 5550 Mayflower Hill, Waterville, ME 04901, United States.

E-mail address: christopher.soto@colby.edu (C.J. Soto).

2013; Soto, John, Gosling, & Potter, 2008) and associations with external criteria (Danner, Aichholzer, & Rammstedt, 2015). By including an equal number of true-keyed and false-keyed items on each domain and facet scale, the BFI-2 automatically controls individual differences in acquiescence at the scale level. This balanced item content also allows researchers to easily control acquiescence in item-level analyses, either by estimating latent variable models that include an acquiescence method factor (e.g., Aichholzer, 2014; Soto & John, *in press*), or through simple within-person centering: subtracting an individual's overall mean response across the full set of 60 BFI-2 items from each of their individual item responses. (However, note that within-person centering can sometimes introduce other psychometric problems; Baron, 1996.)

Third, the BFI-2 is easy to understand. Its items are short phrases that elaborate on a trait-descriptive adjective (e.g., persistent) by adding a synonym, definition, or context (e.g., "Is persistent, works until the task is finished."). These phrased items retain the simplicity and brevity of adjective ratings, while addressing the limitation that individual trait adjectives often have ambiguous or multiple meanings (Goldberg & Kilkowski, 1985). Finally, the BFI-2 is efficient. Its 60 items can be completed in approximately 5–10 min, whereas many broadband personality measures include hundreds of items and can take an hour or more to administer.

The full BFI-2's reasonably short completion time makes it appropriate for many basic and applied research contexts. However, there are some circumstances in which administering the full set of 60 items may not be feasible, and an even shorter measure is needed. For example, some large-scale surveys—such as the British Household Panel Survey (Taylor, Brice, Buck, & Prentice-Lane, 2010), the German Socio-Economic Panel study (Wagner, Frick, & Schupp, 2007), and the Household Income and Labour Dynamics in Australia survey (Summerfield et al., 2015)—are designed to measure many dozens of personal and environmental characteristics as efficiently as possible. When assessing each participant, such surveys may only be able to devote a minute or two to assessing personality traits. Another circumstance concerns within-subjects designs that ask participants to complete the same personality measure multiple times. For example, a single participant may be asked to rate their own personality in several different contexts (Wood & Roberts, 2006), or to rate several other participants in a round-robin design (Srivastava, Guglielmo, & Beer, 2010). In such situations, very brief measures may be needed to prevent participant fatigue, frustration, and careless responding. Finally, some laboratory studies may wish to briefly assess personality traits while still reserving as much time as possible for experimental manipulations and direct behavioral observation.

1.3. A bottom-up strategy for developing the BFI-2 short forms

Given their possible value and most likely uses, we developed the BFI-2-S and BFI-2-XS with two key goals in mind. First, we wanted the short forms to coherently assess each Big Five domain and clearly differentiate between the domains, thereby retaining the BFI-2's clear Big Five structure. Second, we also wanted the short forms to adequately represent each domain's considerable bandwidth—rather than narrowing the range of personality content assessed—in order to maintain the BFI-2's descriptive and predictive breadth. To help balance these two goals, we used a bottom-up approach to scale construction organized around the 15 BFI-2 facets. Specifically, we constructed the BFI-2-XS by selecting a single item to represent each facet, and then constructed the BFI-2-S by adding a second item per facet.

Because the BFI-2 facets have a clear Big Five structure (Soto & John, *in press*), we expected that this strategy would provide the

short forms with a similarly robust domain-level structure. And because same-domain BFI-2 facets can be meaningfully distinguished from each other (Soto & John, *in press*), we also expected that this strategy would preserve a suitably broad range of content within each domain. Furthermore, selecting an item set that equally represents each BFI-2 facet within the Big Five domains raises the possibility that the short forms, like the full measure, might prove useful for assessing personality traits hierarchically. While validating the BFI-2-S and BFI-2-XS, we therefore investigated whether these short forms should only be used to assess the Big Five domains, or whether they are also appropriate for examining facet-level traits.

Despite its strengths, we expected that our bottom-up approach to constructing the BFI-2 short forms would also have some drawbacks. Perhaps most notably, compared with alternative strategies focused on maximizing internal consistency within each Big Five domain (e.g., by selecting items with especially high content overlap, high inter-item correlations, or high domain-level factor loadings), we expected that representing each BFI-2 facet equally might result in relatively low internal consistency reliability for some of the six-item BFI-2-S and (especially) three-item BFI-2-XS domain scales. However, reviews of the psychometric literature have noted that content breadth is generally more important than internal consistency for enhancing the validity of brief measures (Smith, McCarthy, & Anderson, 2000; see also John & Soto, 2007; Stanton, Sinar, Balzer, & Smith, 2002). Thus, prioritizing content validity over internal consistency should help the BFI-2 short forms retain as much of the full measure's validity as possible (cf. Gosling, Rentfrow, & Swann, 2003; Rammstedt & John, 2007).

1.4. Overview of the present research

In sum, the present research was conducted to develop two short forms of the BFI-2—the 30-item BFI-2-S and the 15-item BFI-2-XS—and to address two key research questions. First, to what extent do the BFI-2 short forms retain the reliability and validity of the full measure? Second, is it appropriate to use the BFI-2 short forms as hierarchical personality measures? In other words, should the BFI-2-S and BFI-2-XS only be used to assess personality at the level of the Big Five domains, or are they also appropriate for examining facet-level traits?

2. Study 1

Study 1 had two main goals. The first was to select items for the BFI-2-S and BFI-2-XS, using a joint rational-empirical approach to scale construction. The second was to examine the two short forms' basic measurement properties, using data from three samples. Ideally, the short forms would converge strongly with the full BFI-2 domain scales, demonstrate adequate reliability, and retain a clear Big Five structure.

2.1. Method

2.1.1. Participants and procedure

Study 1 analyzed data from three item selection samples: an Internet sample, a university sample, and a college sample. As described below, data from some of these participants were previously analyzed to validate the full BFI-2 (Soto & John, *in press*, Study 3). However, none of the present data overlapped with those used to select items for the full measure (Soto & John, *in press*, Study 2).

2.1.1.1. Internet selection sample. Participants in this sample were 1000 adult visitors to a personality test website (50% male, 50%

female) who completed the BFI-2 online in exchange for automatically generated feedback. Their median age was 24 years old, with 77% between ages 18 and 35. Regarding ethnicity, 66% described their ethnic identification as White/Caucasian, 7% as Asian/Asian-American, 7% as Hispanic/Latino, 6% as Black/African-American, 1% as Native American/American Indian, 4% as another ethnicity, and 5% as multiple ethnicities, with 5% not reporting ethnicity. Most participants (82%) were residents of the United States, with smaller numbers residing in the United Kingdom (9%), Canada (7%), and Australia or New Zealand (3%). These participants previously constituted the Internet sample in Study 3 of Soto and John (in press).

2.1.1.2. University selection sample. Participants in this sample were 784 students (68% female, 32% male) enrolled in introductory psychology and business courses at the University of California, Berkeley, who completed the BFI-2 online in exchange for partial fulfillment of a course requirement. Their median age was 21 years old, with 95% between ages 18 and 25; 46% described their primary ethnic identification as Asian/Asian-American, 35% as White/Caucasian, 13% as Hispanic/Latino, 2% as Black/African-American, and 1% as another ethnicity, with 3% not reporting ethnicity. A subsample of 137 participants also completed the BFI-2 a second time, with an average retest interval of approximately two months; 110 of these participants were also part of the student sample analyzed in Study 3 of Soto and John (in press).

2.1.1.3. College selection sample. Participants in this sample were 318 students at Colby College (63% female, 29% male, 1% another gender, 7% did not report gender) who completed the BFI-2 online in exchange for a chance to win one of several gift cards. Their median age was 20 years old, with 99% between ages 18 and 22; 70% described their ethnic identification as White/Caucasian, 9% as Asian/Asian American, 2% as Hispanic/Latino, 2% as Black/African American, 1% as another ethnicity, and 9% as multiple ethnicities, with 7% not reporting ethnicity. A subsample of 121 participants also completed the BFI-2 a second time, with an average retest interval of approximately three months.

2.1.2. The Big Five Inventory-2

The BFI-2 (Soto & John, in press) is a hierarchical measure of the Big Five personality domains and 15 more-specific facet traits. Its 60 items are short, descriptive phrases with the common item stem “I am someone who. . .” (e.g., “Is outgoing, sociable,” “Tends to be quiet”). Respondents rate each item using a 5-point scale ranging from *disagree strongly* to *agree strongly*. Soto and John (in press) provided evidence for the reliability, structure, and validity of the BFI-2 domain and facet scales. Alpha reliabilities of the 12-item domain scales averaged 0.86 in each of the three present samples, with a total range of 0.81 to 0.90 across samples. Alphas of the four-item facet scales averaged 0.75 in each sample, with a total range of 0.59 to 0.86.

2.2. Results and discussion

2.2.1. Developing the BFI-2-XS

The first goal of Study 1 was to develop short forms that would accurately represent the content and structure of the full BFI-2. With this goal in mind, we constructed the 15-item BFI-2-XS by selecting a single item from each of the 15 BFI-2 facet scales, using a combination of empirical and rational criteria. The first criterion was each individual item’s correlation with its total facet scale. The second criterion was each item’s standardized loading on its facet factor in a bifactor confirmatory factor analysis (CFA) model of the facet scale. This CFA model allowed each item to load on a substantive facet factor and an acquiescence method factor; to ensure that

the latter factor would represent acquiescence, all items (including both true-keyed and false-keyed items) were constrained to load equally on it, and the acquiescence factor was not allowed to correlate with the facet factor (cf. Soto & John, in press, Fig. 2). The third criterion was an item response theory (IRT) estimate of the total information that each item provided about its facet trait, aggregated across the range of three standard deviations below to three standard deviations above the trait’s mean level. The fourth criterion was the two authors’ conceptual judgments regarding the extent to which each item’s content represents the overall meaning of its facet scale. The fifth criterion was each item’s retest reliability in the university and college samples. The sixth criterion was each item’s pattern of loadings in an exploratory factor analysis (EFA) and a principal components analysis (PCA) of the preliminary BFI-2-XS that extracted and varimax-rotated five dimensions. The seventh and final criterion was to ensure that the BFI-2-XS included both true-keyed and false-keyed items within each Big Five domain. Together, these seven criteria were designed to identify a set of 15 items that (a) represent the content and overall meaning of each BFI-2 domain (criteria 1, 2, 3, and 4), (b) are rated reliably (criterion 5), (c) have a clear Big Five structure (criterion 6), and (d) minimize the influence of acquiescent responding (criterion 7).

For seven of the 15 facets, these criteria consistently suggested one particular item for selection. For the remaining facets, they identified two (for Energy Level, Compassion, Organization, Productiveness, Depression, and Intellectual Curiosity) or three (for Respectfulness and Creative Imagination) viable items. In these cases, we finalized our selection through discussion of each item’s empirical and conceptual strengths and weaknesses. For example, the Organization items “Tends to be disorganized” and “Keeps things neat and tidy” had similar empirical properties. After discussion, we selected the former item due to its broader psychological meaning, as compared with the latter item’s focus on physical neatness. The complete set of 15 items selected for the BFI-2-XS is provided in Appendix A.

2.2.2. Developing the BFI-2-S

We constructed the 30-item BFI-2-S by supplementing the 15-item BFI-2-XS with one additional item from each of the BFI-2 facet scales.¹ As with the BFI-2-XS, we selected these supplemental items using a set of empirical and rational criteria. Our first criterion was the correlation of each two-item facet composite (the BFI-2-XS item plus the supplemental item) with its total facet scale. The second criterion was the total information provided by each two-item composite about its facet trait, as estimated from an IRT analysis of the complete facet scale. The third criterion was each two-item composite’s retest reliability in the undergraduate and college samples. The fourth criterion was the inter-item correlation within each two-item composite; ideally, this correlation would be strong enough to indicate consistent responding but not so strong as to indicate excessive redundancy. The fifth criterion was the two authors’ conceptual judgments of breadth vs. redundancy in item content (with a preference for greater breadth). The sixth criterion was each item’s pattern of loadings in an EFA and a PCA of the preliminary BFI-2-S that extracted and varimax-rotated five dimensions. The seventh and final criterion was to ensure that the BFI-2-S included one true-keyed item and one false-keyed item for each facet. These criteria were designed to select a set of 30 items that (a) represent the content and meaning of each BFI-2 domain (criteria 1 and 2), (b) are

¹ An alternative strategy would have been to select item pairs for the BFI-2-S independently, without requiring that they include the BFI-2-XS items. Preliminary analyses suggested that this would have produced no more than trivial differences in our final item selections. We therefore decided to maximize comparability between the two short forms by nesting the BFI-2-XS items within the BFI-2-S.

reported reliably (criterion 3), (c) avoid excessive redundancy between items (criteria 4 and 5), (d) have a clear Big Five structure (criterion 6), and (e) minimize the influence of acquiescence (criterion 7).

For 10 of the 15 facets, these criteria consistently suggested one particular item for selection. For the remaining facets (Assertiveness, Energy Level, Depression, Emotional Volatility, and Creative Imagination), they identified two viable items. We resolved these cases through discussion. For example, the Energy Level items “Is less active than other people” and “Rarely feels excited or eager” were both viable supplements to the BFI-2-XS item “Is full of energy.” After discussion, we selected the former item due to its higher retest reliability and weaker secondary loading on Agreeableness. The complete set of 30 items selected for the BFI-2-S is provided in [Appendix A](#).

2.2.3. A preliminary examination of the short forms' measurement properties

The second goal of Study 1 was to briefly examine the BFI-2 short forms' basic measurement properties. To do this, we examined the part-whole correlations (i.e., the correlation of each short-form scale with the corresponding full BFI-2 scale; cf. [Rammstedt & John, 2007](#)), reliabilities, and multidimensional structure of the BFI-2-S and BFI-2-XS in the three selection samples. Part-whole correlations of the six-item BFI-2-S domain scales with the full, 12-item BFI-2 domain scales averaged either 0.95 or 0.96 in each sample (total range across the three samples = 0.93–0.97), and correlations of the three-item BFI-2-XS domain scales with the full BFI-2 domain scales averaged 0.90 in each sample (total range = 0.85–0.93). These results suggest that our empirical and rational item selection criteria successfully produced short forms that adequately represent the content and overall meaning of the BFI-2 domains, with the BFI-2-S providing (as expected) better representation than the BFI-2-XS.

Alpha reliabilities of the BFI-2-S domain scales averaged 0.77 or 0.78 in each sample (total range = 0.73–0.83); these scales' retest reliabilities averaged 0.76 in the university sample (range = 0.69–0.83) and 0.83 in the college sample (range = 0.77–0.88). Alphas of the BFI-2-XS domain scales averaged between 0.61 and 0.63 in each sample (total range = 0.51–0.72); these scales' retest reliabilities averaged 0.70 in the university sample (range = 0.60–0.80) and 0.76 in the college sample (range = 0.71–0.80). These results suggest that both short forms are adequately reliable, with the BFI-2-S providing greater reliability than the BFI-2-XS. They also suggest that the BFI-2-XS scales tend to have greater retest reliability than alpha reliability, a pattern typical for brief scales that pri-

oritize content validity over internal consistency (cf. [Gosling et al., 2003](#)).

To examine the multidimensional structure of the BFI-2-S and BFI-2-XS, we submitted each short form's items to a random intercept EFA that extracted and varimax-rotated five factors, using maximum likelihood estimation. Random intercept EFA ([Aichholzer, 2014](#)) is a procedure for examining the multidimensional structure of an item set while using a method factor to model individual differences in acquiescent responding, thereby minimizing the negative structural effects of acquiescence variance ([Rammstedt & Farmer, 2013](#); [Soto et al., 2008](#)). For both short forms, each item had its strongest loading on the expected factor in all three samples. For the BFI-2-S, the absolute primary loadings averaged 0.59 or 0.60 in each sample, whereas the absolute secondary loadings averaged only 0.09 or 0.10. For the BFI-2-XS, the absolute primary loadings averaged 0.59 to 0.61 in each sample, whereas the absolute secondary loadings averaged only 0.10. For both short forms, all congruence coefficients comparing pairs of corresponding factors across samples were at least 0.96. These results suggest that both the BFI-2-S and BFI-2-XS have a clear Big Five structure.

3. Study 2

Study 1 selected items for the 15-item BFI-2-XS and the 30-item BFI-2-S, and provided preliminary evidence regarding these short forms' part-whole convergence, reliability, and multidimensional structure in three samples. However, the measurement properties observed in Study 1 were potentially biased by the fact that we selected items for the short forms based partly on exploratory analyses of these same samples. Therefore, one major goal of Study 2 was to more thoroughly examine the reliability and validity of the BFI-2-S and BFI-2-XS using data from two independent validation samples. Our second major goal was to test whether the BFI-2 short forms should only be used to assess personality at the level of the broad Big Five domains, or whether they are also appropriate for examining more-specific, facet-level traits within each domain.

3.1. Method

3.1.1. Participants and procedure

Study 2 analyzed data from two validation samples: an Internet sample and a university sample. As described below, some of the present data were previously analyzed to validate the full BFI-2 ([Soto & John, in press, Study 3](#)). However, none of these data were used to select items for the full measure ([Soto & John, in press,](#)

Table 1
Summary and comparison of measurement properties for the BFI-2, BFI-2-S, and BFI-2-XS domain scales (Study 2).

Domain scale	Part-whole correlations				Alpha reliabilities						Self-peer correlations		
	BFI-2-S		BFI-2-XS		BFI-2		BFI-2-S		BFI-2-XS		BFI-2	BFI-2-S	BFI-2-XS
	Internet	Univers.	Internet	Univers.	Internet	Univers.	Internet	Univers.	Internet	Univers.	Univers.	Univers.	Univers.
Extraversion	0.95	0.96	0.90	0.92	0.86	0.87	0.77	0.78	0.63	0.66	0.69	0.66	0.59
Agreeableness	0.94	0.94	0.86	0.87	0.82	0.83	0.75	0.75	0.55	0.49	0.47	0.49	0.48
Conscientiousness	0.96	0.95	0.89	0.88	0.88	0.85	0.78	0.73	0.61	0.55	0.53	0.47	0.45
Negative Emotionality	0.97	0.97	0.92	0.92	0.90	0.86	0.84	0.82	0.73	0.69	0.62	0.59	0.56
Open-Mindedness	0.95	0.95	0.88	0.90	0.84	0.84	0.74	0.77	0.57	0.58	0.47	0.46	0.42
Mean coefficient	0.95	0.95	0.89	0.90	0.86	0.85	0.78	0.77	0.62	0.59	0.55	0.53	0.50
Proportion of BFI-2 domain scale variance retained	0.91	0.91	0.79	0.80	—	—	0.90	0.91	0.72	0.70	—	0.93	0.82

Note. Internet = Internet validation sample ($N = 2000$). Univers. = University validation sample ($N = 423$; for self-peer correlations, $N = 230$). All correlations are statistically significant at $p < 0.001$. Proportion of BFI-2 domain scale variance retained equals the mean squared part-whole correlation, the ratio of the mean alpha reliability for a short form compared to the full BFI-2, or the ratio of the mean squared self-peer correlation for a short form compared to the full BFI-2.

Table 2
Loadings of the BFI-2-S items in a random intercept EFA with Varimax rotation (Study 2).

Domain, facet, and item text	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality	Open-Mindedness
Extraversion					
<i>Sociability</i>					
Is outgoing, sociable.	0.69/0.76	0.21/0.10	0.04/0.02	−0.10/−0.08	0.02/0.10
Tends to be quiet.	− 0.68/−0.72	−0.01/0.04	0.08/0.07	0.02/−0.02	0.03/0.01
<i>Assertiveness</i>					
Is dominant, acts as a leader.	0.61/0.59	−0.19/−0.16	0.20/0.12	−0.08/−0.08	0.17/0.11
Prefers to have others take charge.	− 0.50/−0.43	0.10/0.14	−0.18/−0.17	0.14/0.15	−0.20/−0.21
<i>Energy Level</i>					
Is full of energy.	0.57/0.64	0.18/0.14	0.15/0.06	−0.24/−0.14	0.03/−0.09
Is less active than other people.	− 0.39/−0.44	−0.09/−0.04	−0.23/−0.23	0.24/0.22	−0.14/−0.16
Agreeableness					
<i>Compassion</i>					
Is compassionate, has a soft heart.	0.12/−0.03	0.56/0.54	0.02/0.13	0.17/0.14	0.13/0.23
Can be cold and uncaring.	−0.11/−0.10	− 0.66/−0.71	−0.09/−0.05	0.10/0.05	−0.02/−0.07
<i>Respectfulness</i>					
Is respectful, treats others with respect.	0.06/−0.10	0.52/0.50	0.16/0.23	−0.12/−0.14	0.11/0.17
Is sometimes rude to others.	0.11/0.15	− 0.65/−0.56	−0.12/−0.05	0.21/0.28	−0.06/−0.08
<i>Trust</i>					
Assumes the best about people.	0.13/0.19	0.51/0.56	−0.01/0.02	−0.13/−0.20	0.01/−0.10
Tends to find fault with others.	−0.04/−0.06	− 0.52/−0.57	−0.05/0.01	0.28/0.22	0.00/0.01
Conscientiousness					
<i>Organization</i>					
Keeps things neat and tidy.	0.08/0.04	0.04/0.03	0.72/0.69	−0.04/−0.01	−0.07/0.00
Tends to be disorganized.	−0.04/−0.01	−0.05/−0.06	− 0.81/−0.80	0.09/0.12	0.04/0.01
<i>Productiveness</i>					
Is persistent, works until the task is finished.	0.12/0.11	0.08/0.03	0.52/0.43	−0.15/−0.12	0.10/0.10
Has difficulty getting started on tasks.	−0.16/−0.11	−0.04/−0.04	− 0.51/−0.44	0.21/0.22	−0.03/0.08
<i>Responsibility</i>					
Is reliable, can always be counted on.	0.12/0.07	0.25/0.26	0.43/0.40	−0.19/−0.18	0.06/0.20
Can be somewhat careless.	0.02/−0.09	−0.25/−0.19	− 0.49/−0.43	0.14/0.10	−0.05/−0.14
Negative Emotionality					
<i>Anxiety</i>					
Worries a lot.	−0.14/−0.19	−0.03/−0.07	−0.03/−0.04	0.65/0.60	−0.02/−0.07
Is relaxed, handles stress well.	0.12/0.11	0.09/0.00	0.07/0.07	− 0.69/−0.65	0.05/0.02
<i>Depression</i>					
Tends to feel depressed, blue.	−0.26/−0.28	−0.16/−0.23	−0.18/−0.22	0.61/0.51	0.02/0.05
Feels secure, comfortable with self.	0.31/0.35	0.08/0.18	0.21/0.19	− 0.52/−0.47	0.06/−0.01
<i>Emotional Volatility</i>					
Is temperamental, gets emotional easily.	0.08/0.05	−0.06/−0.12	−0.08/−0.07	0.70/0.70	−0.05/−0.06
Is emotionally stable, not easily upset.	0.04/0.04	0.08/0.07	0.07/0.07	− 0.78/−0.81	0.02/−0.03
Open-Mindedness					
<i>Aesthetic Sensitivity</i>					
Is fascinated by art, music, or literature.	−0.01/−0.09	0.06/0.09	−0.05/0.01	0.07/0.11	0.53/0.61
Has few artistic interests.	0.00/0.04	−0.01/−0.10	0.05/0.00	0.00/−0.05	− 0.57/−0.65
<i>Intellectual Curiosity</i>					
Is complex, a deep thinker.	0.04/0.04	−0.04/−0.02	0.04/0.15	0.13/0.05	0.48/0.46
Has little interest in abstract ideas.	0.00/−0.03	−0.06/−0.05	0.09/−0.03	0.11/0.04	− 0.58/−0.62
<i>Creative Imagination</i>					
Is original, comes up with new ideas.	0.22/0.27	0.09/0.02	0.07/0.00	−0.11/−0.16	0.58/0.51
Has little creativity.	−0.08/−0.15	−0.08/−0.04	0.01/0.00	0.07/0.08	− 0.71/−0.72

Note. BFI-2 items copyright 2015 by Oliver P. John and Christopher J. Soto. Reprinted with permission. EFA = Exploratory factor analysis. Random intercept EFA constrains each item to load on an acquiescence method factor, in addition to the substantive factors. Loadings left of the forward slash are from the Internet validation sample ($N = 2000$). Loadings right of the forward slash are from the university validation sample ($N = 423$). The average loading on the acquiescence method factor was 0.14 (range = 0.13–0.23) in the Internet sample and 0.15 (range = 0.12–0.24) in the university sample. Absolute loadings ≥ 0.30 are bolded.

Study 2) or the short forms (Study 1, above). Therefore, the present study can provide unbiased estimates of the BFI-2 short forms' measurement properties.

3.1.1.1. Internet validation sample. Participants in this sample were 2000 adults (50% male, 50% female) recruited and assessed using the same procedure as the Internet selection sample analyzed in

Study 1. The median age of participants in the Internet validation sample was 24 years old, with 77% between ages 18 and 35. Regarding ethnicity, 64% described themselves as White/Caucasian, 8% as Black/African-American, 7% as Asian/Asian-American, 7% as Hispanic/Latino, 1% as Native American/American Indian, 4% as another ethnicity, and 4% as multiple ethnicities, with 5% not reporting ethnicity. Participants were residents of the

United States (80%), the United Kingdom (8%), Canada (7%), or Australia or New Zealand (5%).

3.1.1.2. University validation sample. Participants in this sample were 423 students (66% female, 31% male, 3% did not report gender) enrolled in introductory psychology courses at the University of California, Berkeley, who completed the BFI-2 online in exchange for partial fulfillment of a course requirement. Their median age was 21 years old, with 90% between ages 18 and 25. Regarding ethnicity, 50% described themselves as Asian/Asian-American, 24% as White/Caucasian, 12% as Hispanic/Latino, 2% as Black/African-American, and 8% as another ethnicity, with 4% not reporting ethnicity. Most of these participants (360) were also members of the student sample analyzed in Study 3 of Soto and John (in press).

3.1.2. Measures

All participants in both validation samples completed the full BFI-2 item set, which was then used to score the BFI-2-S and BFI-2-XS. Some members of the university sample were also assessed using a set of self-reported and peer-reported criteria, described below. (For further details about these measures, see Soto & John, in press.)

3.1.2.1. Behavioral self-reports. Approximately two weeks after completing the BFI-2, 392 members of the university validation sample described their behavior during the previous six months using a set of 80 items (Bardi & Schwartz, 2003). Each item was rated on a 5-point frequency scale ranging from *never* to *all the time*. Following the recommendation of the original authors (Bardi & Schwartz, 2003), each respondent's complete set of ratings was within-person centered to remove the substantial individual differences in acquiescent responding typically observed for this measure. The centered items were then aggregated into scales corresponding with the 10 values of the Schwartz values circumplex: conformity, tradition, benevolence, power, universalism, hedonism, security, stimulation, achievement, and self-direction. As in previous research, after within-person centering, the alpha reliabilities of these scales varied considerably, averaging 0.48 (range = 0.27–0.72) (cf. Bardi & Schwartz, 2003; Pozzebon & Ashton, 2009).

3.1.2.2. Psychological well-being scales. Approximately two weeks before completing the BFI-2, 185 members of the university validation sample completed the Psychological Well-Being Scales (Ryff, 1989), which assess six aspects of well-being: positive relations with others, purpose in life, environmental mastery, self-acceptance, autonomy, and personal growth. These scales include a total of 84 items that respondents rate on a 5-point agreement scale. In this sample, the scales' alpha reliabilities averaged 0.86 (range = 0.84–0.90).

3.1.2.3. Peer-reports. Approximately two months after completing the BFI-2, 230 members of the university validation sample were rated by a knowledgeable peer. Most peers were friends (62%) or romantic partners (28%). Each peer rated the target participant on the full BFI-2, and on a set of additional items assessing four criteria: social connectedness (4 items), likability (2 items), stress resistance (5 items), and positive affect (2 items). All items were rated on a 5-point agreement scale. Alpha reliabilities for the BFI-2 peer-reports averaged 0.86 for the domain scales (range = 0.84–0.90) and 0.75 for the facet scales (range = 0.63–0.84). Alphas for the additional peer-reported criteria were 0.75 for social connectedness, 0.76 for likability, 0.79 for stress resistance, and 0.66 for positive affect.

3.2. Results and discussion

3.2.1. Domain-level measurement properties of the BFI-2 short forms

Table 1 presents part-whole correlations, alpha reliabilities, and self-peer agreement correlations for the BFI-2, BFI-2-S, and BFI-2-XS domain scales. The leftmost part of this table shows that, in each validation sample, part-whole correlations for the BFI-2-S averaged 0.95 (total range across the two samples = 0.94–0.97), and correlations for the BFI-2-XS averaged 0.89 or 0.90 (total range = 0.86–0.92). Closely replicating the preliminary findings from Study 1, these results indicate that the BFI-2-S and BFI-2-XS capture approximately 91% and 80%, respectively, of the total variance in the full BFI-2 domain scales.

The middle part of Table 1 shows that, in each sample, alpha reliabilities of the full BFI-2 domain scales averaged 0.85 or 0.86 (total range = 0.82–0.90), alphas of the BFI-2-S domains averaged 0.77 or 0.78 (total range = 0.73–0.84), and alphas of the BFI-2-XS domains averaged 0.59 or 0.62 (total range = 0.49–0.73). Again replicating the preliminary findings from Study 1, these results indicate a moderate decrease in internal consistency from the full BFI-2 to the BFI-2-S (which retained approximately 91% of the full measure's internal consistency), and a more substantial decrease to the BFI-2-XS (which retained approximately 71% of the full measure's internal consistency). As noted above, these decreases were an expected consequence of our strategy for developing the BFI-2 short forms, which prioritized content validity over internal consistency (cf. Gosling et al., 2003; Smith et al., 2000; Stanton et al., 2002).

Finally, the rightmost part of Table 1 shows that, in the university sample, self-peer agreement correlations averaged 0.55 for the full BFI-2 domain scales (range = 0.47–0.69), 0.53 for the BFI-2-S domains (range = 0.46–0.66), and 0.50 for the BFI-2-XS domains (range = 0.42–0.59). These results indicate that the BFI-2-S and BFI-2-XS retain approximately 93% and 82%, respectively, of the full BFI-2 domain scales' self-peer agreement.

3.2.2. Domain-level structure of the BFI-2 short forms

To examine the domain-level structure of the BFI-2-S and BFI-2-XS, we submitted each short form's items to a random intercept EFA that extracted and varimax-rotated five factors. The factor loadings from these analyses are presented in Tables 2 and 3, respectively. As these tables show, for both short forms each item had its strongest loading on the expected factor in both validation samples. For the BFI-2-S, in each sample the absolute primary loadings averaged 0.58 or 0.59 (total range = 0.39–0.81), whereas the absolute secondary loadings averaged only 0.09 or 0.10 (total range = 0.00–0.35). For the BFI-2-XS, in each sample the absolute primary loadings averaged 0.57 or 0.59 (total range = 0.40–0.76), whereas the absolute secondary loadings averaged only 0.09 or 0.11 (total range = 0.00–0.30). Supporting the utility of modeling acquiescence variance, loadings on the acquiescence method factor were generally small but not trivial, averaging 0.14 or 0.15 in each sample for both short forms (total range = 0.12–0.24). When we examined replication of the factor loadings across the two validation samples, the congruence coefficients between corresponding factors were all at least 0.97 ($M = 0.98$) for the BFI-2-S, and at least 0.95 ($M = 0.97$) for the BFI-2-XS².

² Much previous personality research has used PCAs rather than EFAs to examine multidimensional structure. Other research has used latent variable models that do not impose orthogonality constraints between the Big Five. Therefore, in addition to the analyses reported in Tables 2 and 3, we also conducted PCAs with varimax rotation and random intercept EFAs with direct oblimin rotation ($\gamma = 0$). The results of these analyses are presented in the Online Supplementary Material, Tables S1–S4. Overall, their results were very similar to those reported in Tables 2 and 3. In each analysis, all items had their strongest loading on the expected dimension, and all congruence coefficients with the corresponding factors reported in Tables 2 and 3 were at least .97. Thus, the present results were quite robust across alternative methods.

Table 3
Loadings of the BFI-2-XS items in a random intercept EFA with Varimax rotation (Study 2).

Domain and item text	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality	Open-Mindedness
<i>Extraversion</i>					
Tends to be quiet.	−0.66/−0.76	−0.01/0.01	0.07/0.07	0.08/0.02	0.02/−0.03
Is dominant, acts as a leader.	0.59/0.58	−0.14/−0.15	0.22/0.14	−0.07/−0.09	0.15/0.12
Is full of energy.	0.55/0.59	0.20/0.17	0.17/0.09	−0.28/−0.19	0.02/−0.08
<i>Agreeableness</i>					
Is compassionate, has a soft heart.	0.11/−0.06	0.58/0.49	0.04/0.20	0.13/0.16	0.10/0.22
Is sometimes rude to others.	0.16/0.19	−0.58/−0.49	−0.13/−0.02	0.23/0.30	−0.08/−0.08
Assumes the best about people.	0.09/0.17	0.51/0.55	0.00/0.04	−0.16/−0.24	0.04/−0.07
<i>Conscientiousness</i>					
Tends to be disorganized.	−0.03/0.00	−0.04/−0.05	−0.73/−0.70	0.10/0.11	0.04/−0.01
Has difficulty getting started on tasks.	−0.15/−0.08	−0.03/−0.04	−0.54/−0.45	0.20/0.24	0.01/0.10
Is reliable, can always be counted on.	0.11/0.01	0.22/0.21	0.45/0.40	−0.18/−0.20	0.03/0.25
<i>Negative Emotionality</i>					
Worries a lot.	−0.10/−0.13	0.00/−0.07	−0.03/−0.03	0.66/0.59	−0.05/−0.05
Tends to feel depressed, blue.	−0.21/−0.23	−0.10/−0.19	−0.19/−0.26	0.67/0.55	0.01/0.05
Is emotionally stable, not easily upset.	0.01/0.02	0.04/0.04	0.10/0.11	−0.71/−0.74	0.05/0.02
<i>Open-Mindedness</i>					
Is fascinated by art, music, or literature.	0.01/−0.04	−0.05/0.00	0.09/−0.02	0.08/0.04	−0.64/−0.72
Has little interest in abstract ideas.	0.02/−0.08	0.09/0.09	−0.03/0.01	0.09/0.14	0.53/0.55
Is original, comes up with new ideas.	0.22/0.24	0.10/0.08	0.10/−0.04	−0.11/−0.17	0.50/0.46

Note. BFI-2 items copyright 2015 by Oliver P. John and Christopher J. Soto. Reprinted with permission. EFA = Exploratory factor analysis. Random intercept EFA constrains each item to load on an acquiescence method factor, in addition to the substantive factors. Loadings left of the forward slash are from the Internet validation sample ($N = 2000$). Loadings right of the forward slash are from the university validation sample ($N = 423$). The average loading on the acquiescence method factor was 0.14 (range = 0.12–0.18) in the Internet sample and 0.15 (range = 0.12–0.21) in the university sample. Absolute loadings ≥ 0.30 are bolded.

Replicating the preliminary findings from Study 1, these results indicate that both BFI-2 short forms have a clear Big Five structure.

3.2.3. Domain-level external validity of the BFI-2 short forms

To examine the BFI-2 short forms' external validity, we first computed correlations of the BFI-2, BFI-2-S, and BFI-2-XS domain scales with the set of 20 self-reported and peer-reported criteria assessed in the university validation sample; we then regressed each criterion on each set of domain scales. The resulting correlations and standardized regression coefficients are presented in Table 4. As this table shows, the pattern of criterion associations was very similar across the three BFI-2 forms. Column-vector correlations comparing the pattern of domain-criterion correlations for the full BFI-2 against the BFI-2-S were at least 0.99 for each of the 20 criteria, and the corresponding column-vector correlations comparing the patterns of standardized regression coefficients were all at least 0.92 ($M = 0.99$). When comparing the BFI-2 with the BFI-2-XS, column-vector correlations for the pattern of domain-criterion correlations were all at least 0.97 ($M = 0.99$), and those for the pattern of regression coefficients were all at least 0.88 ($M = 0.97$).

Beyond these similar patterns of criterion associations, how well do the BFI-2 short forms retain the full domain scales' overall level of predictive power? To address this question, we compared the total proportion of variance in each criterion variable explained by the BFI-2, BFI-2-S, and BFI-2-XS domain scales, respectively; these proportions are presented in Table 5. As this table shows, the BFI-2-S retained approximately 93% (mean $R^2 = 0.236$), and the BFI-2-XS approximately 84% (mean $R^2 = 0.214$), of the full BFI-2's predictive power (mean $R^2 = 0.254$). Inspection of the individual criteria revealed that this pattern of decreasing predictive power with decreasing scale length generalized across the behavioral, psychological, and peer-reported criteria. Taken together, these results indicate that the BFI-2 short forms capture the full measure's domain-level pattern of substantive associations with external criteria quite well, but that using a short form instead of the full measure entails a loss of overall predictive power.

3.2.4. Facet-level external validity of the BFI-2 short forms

The results presented thus far converge in indicating that the BFI-2-S and BFI-2-XS efficiently measure the Big Five domains. They also indicate that developing these short forms using a strategy that prioritized content breadth over internal consistency helped them retain much of the full BFI-2's domain-level validity. Are the short forms also appropriate for examining facet-level traits? To address this question, we repeated the correlation and regression analyses of external validity, just described, at the facet level.

As with the domain-level analyses, the pattern of facet-criterion correlations was similar across the three BFI-2 forms. Column-vector correlations comparing the pattern of criterion correlations for the four-item BFI-2 facet scales against the two-item BFI-2-S scales were at least 0.95 ($M = 0.98$) for each of the 20 criteria, and those comparing the BFI-2 facet scales against the single BFI-2-XS items were all at least 0.86 ($M = 0.95$). The pattern of facet-level standardized regression coefficients was also reasonably similar between the full BFI-2 and the BFI-2-S for the 10 self-reported behavioral criteria. Column-vector correlations comparing these measures' regression coefficients were at least 0.78 ($M = 0.90$) for each of these 10 criteria. Moreover, averaged across these criteria, the BFI-2-S facets retained approximately 89% (mean $R^2 = 0.215$) of the full BFI-2's predictive power (mean $R^2 = 0.241$).

However, the pattern of facet-level standardized regression coefficients was much less consistent between the BFI-2 and the BFI-2-S for the well-being and peer-reported criteria. Column-vector correlations comparing these measures' regression coefficients averaged only 0.64 (minimum = -0.15) across these 10 criteria. The facet-level regression coefficients were also rather inconsistent between the full BFI-2 and the BFI-2-XS for the complete set of 20 self-reported and peer-reported criteria, with column-vector correlations averaging only 0.75 (minimum = 0.39).

Taken together, this pattern of results suggests three conclusions. First, the BFI-2-S may be useful for examining facet-level associations in reasonably large samples. The self-reported behavioral criteria were assessed in a sample of approximately 400 participants, and the BFI-2 and BFI-2-S produced similar patterns of facet-criterion associations in this sample, although the short form

Table 4
Associations of the BFI-2, BFI-2-S, and BFI-2-XS domain scales with self-reported and peer-reported criteria (Study 2).

Criterion	Correlations					Standardized regression coefficients				
	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality	Open-Mindedness	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality	Open-Mindedness
<i>Self-reported behavioral criteria</i>										
Conformity	-0.34/-0.33/-0.32	0.23/0.21/0.19	0.13/0.13/0.15	0.05/0.06/0.06	-0.15/-0.16/-0.14	-0.35/-0.34/-0.31	0.26/0.22/0.19	0.20/0.20/0.18	0.10/0.08/0.09	-0.15/-0.15/-0.14
Tradition	-0.23/-0.21/-0.20	0.05/0.04/0.06	0.00/-0.01/0.03	-0.11/-0.10/-0.09	-0.16/-0.12/-0.12	-0.27/-0.25/-0.23	0.01/0.02/0.03	0.02/0.00/0.00	-0.18/-0.18/-0.14	-0.12/-0.09/-0.11
Benevolence	0.08/0.06/0.01	0.51/0.46/0.41	0.20/0.19/0.21	-0.11/-0.13/-0.09	0.16/0.12/0.11	0.05/0.02/0.00	0.52/0.44/0.39	0.06/0.09/0.15	0.11/0.05/0.06	0.05/0.06/0.06
Power	0.33/0.33/0.39	-0.51/-0.45/-0.44	-0.12/-0.11/-0.13	0.05/0.06/0.04	-0.20/-0.17/-0.16	0.40/0.41/0.41	-0.49/-0.43/-0.41	-0.07/-0.07/-0.07	-0.04/0.03/0.00	-0.17/-0.17/-0.15
Universalism	-0.01/-0.02/-0.01	0.14/0.15/0.18	-0.05/-0.04/-0.06	0.02/0.00/-0.01	0.14/0.13/0.12	0.00/-0.02/-0.01	0.17/0.16/0.20	-0.11/-0.08/-0.10	0.04/0.02/0.01	0.13/0.13//0.10
Hedonism	-0.06/-0.08/-0.05	-0.21/-0.20/-0.23	-0.38/-0.36/-0.34	0.17/0.16/0.14	-0.07/-0.06/-0.05	0.04/0.01/-0.02	-0.10/-0.11/-0.16	-0.36/-0.33/-0.30	0.01/0.03/-0.01	-0.01/-0.01/-0.02
Stimulation	0.23/0.22/0.21	-0.12/-0.10/-0.11	-0.23/-0.22/-0.23	-0.02/-0.02/-0.02	0.15/0.13/0.11	0.26/0.25/0.21	-0.10/-0.08/-0.08	-0.33/-0.29/-0.26	-0.09/-0.05/-0.07	0.17/0.13/0.11
Security	-0.22/-0.20/-0.20	0.04/0.03/0.03	0.21/0.19/0.17	0.00/0.01/-0.01	-0.15/-0.13/-0.12	-0.27/-0.23/-0.21	0.00/0.00/0.01	0.31/0.27/0.19	0.02/0.02/0.00	-0.14/-0.13/-0.11
Achievement	0.20/0.19/0.17	-0.09/-0.06/-0.04	0.19/0.19/0.16	0.01/0.02/0.02	-0.02/-0.01/-0.02	0.19/0.20/0.18	-0.11/-0.08/-0.06	0.22/0.20/0.19	0.11/0.12/0.10	-0.07/-0.05/-0.03
Self-direction	0.05/0.06/0.00	-0.05/-0.08/-0.07	0.04/0.02/-0.02	-0.05/-0.07/-0.02	0.45/0.42/0.42	-0.07/-0.04/-0.05	-0.18/-0.16/-0.13	0.00/-0.02/-0.02	-0.11/-0.12/-0.06	0.49/0.45/0.43
<i>Self-reported well-being criteria</i>										
Positive relations	0.41/0.40/0.32	0.45/0.46/0.35	0.30/0.26/0.27	-0.39/-0.39/-0.40	0.26/0.24/0.21	0.33/0.32/0.26	0.35/0.37/0.25	0.00/-0.02/0.06	-0.14/-0.16/-0.24	0.09/0.12/0.13
Environmental mastery	0.44/0.45/0.34	0.32/0.29/0.23	0.60/0.58/0.60	-0.59/-0.55/-0.54	0.15/0.13/0.13	0.22/0.24/0.23	0.01/0.05/0.00	0.38/0.38/0.45	-0.37/-0.34/-0.31	0.01/0.00/0.04
Purpose in life	0.42/0.41/0.32	0.30/0.29/0.21	0.53/0.51/0.51	-0.36/-0.37/-0.36	0.25/0.25/0.21	0.26/0.26/0.23	0.08/0.10/0.05	0.37/0.34/0.40	-0.09/-0.14/-0.14	0.10/0.12/0.13
Self-acceptance	0.45/0.48/0.36	0.34/0.33/0.23	0.53/0.50/0.47	-0.57/-0.58/-0.57	0.22/0.21/0.19	0.25/0.28/0.24	0.05/0.10/0.03	0.27/0.25/0.27	-0.36/-0.37/-0.39	0.07/0.09/0.10
Autonomy	0.28/0.26/0.21	0.20/0.16/0.11	0.29/0.27/0.26	-0.31/-0.33/-0.30	0.40/0.37/0.33	0.11/0.10/0.12	-0.02/-0.01/-0.02	0.12/0.11/0.14	-0.22/-0.24/-0.21	0.34/0.32/0.29
Personal growth	0.31/0.31/0.20	0.37/0.32/0.23	0.33/0.31/0.30	-0.22/-0.22/-0.23	0.44/0.38/0.36	0.18/0.22/0.14	0.24/0.23/0.12	0.13/0.12/0.19	0.01/-0.01/-0.07	0.32/0.29/0.31
<i>Peer-reported criteria</i>										
Social connectedness	0.27/0.23/0.23	0.33/0.33/0.29	0.21/0.18/0.17	-0.27/-0.27/-0.27	0.05/0.06/0.07	0.22/0.17/0.18	0.27/0.28/0.23	0.04/0.04/0.06	-0.08/-0.11/-0.13	-0.03/0.02/0.04
Likability/v	0.18/0.16/0.14	0.26/0.24/0.22	0.11/0.07/0.09	-0.10/-0.11/-0.09	0.09/0.11/0.12	0.17/0.14/0.12	0.26/0.23/0.20	0.01/-0.01/0.04	0.06/0.01/0.01	0.02/0.08/0.09
Stress resistance	0.23/0.24/0.19	0.19/0.16/0.13	0.20/0.17/0.18	-0.51/-0.49/-0.45	0.00/-0.01/0.03	0.07/0.08/0.08	0.00/0.00/-0.01	-0.02/-0.02/0.03	-0.50/-0.47/-0.42	0.00/-0.02/0.03
Positive affect	0.37/0.36/0.34	0.14/0.16/0.13	0.24/0.21/0.21	-0.43/-0.43/-0.41	-0.03/-0.01/-0.02	0.26/0.24/0.26	-0.01/0.03/0.02	0.05/0.04/0.07	-0.33/-0.32/-0.31	-0.07/-0.04/-0.03

Note. Positive relations = Positive relations with others. For behavioral criteria, $N = 392$ and absolute correlations ≥ 0.10 are statistically significant at $p < 0.05$. For well-being criteria, $N = 185$ and absolute correlations ≥ 0.15 are statistically significant at $p < 0.05$. For peer-reported criteria, $N = 230$ and absolute correlations ≥ 0.13 are statistically significant at $p < 0.05$. Absolute coefficients ≥ 0.20 are bolded.

Table 5

Proportion of variance in self-reported and peer-reported criteria explained by the BFI-2, BFI-2-S, and BFI-2-XS domain scales (Study 2).

Criterion	BFI-2	BFI-2-S	BFI-2-XS
<i>Self-reported behavioral criteria</i>			
Conformity	0.23	0.21	0.19
Tradition	0.10	0.08	0.08
Benevolence	0.28	0.22	0.19
Power	0.42	0.37	0.37
Universalism	0.05	0.04	0.05
Hedonism	0.16	0.15	0.14
Stimulation	0.17	0.15	0.13
Security	0.15	0.11	0.09
Achievement	0.09	0.08	0.06
Self-direction	0.24	0.21	0.19
Mean	0.19	0.16	0.15
<i>Self-reported well-being criteria</i>			
Positive relations with others	0.37	0.38	0.31
Environmental mastery	0.55	0.53	0.52
Purpose in life	0.39	0.39	0.36
Self-acceptance	0.50	0.52	0.46
Autonomy	0.27	0.25	0.22
Personal growth	0.33	0.29	0.24
Mean	0.40	0.39	0.35
<i>Peer-reported criteria</i>			
Social connectedness	0.18	0.17	0.16
Likability	0.10	0.08	0.07
Stress resistance	0.26	0.24	0.21
Positive affect	0.25	0.24	0.23
Mean	0.20	0.18	0.17
Grand mean	0.25	0.24	0.21
Proportion of BFI-2 domain scale validity retained	—	0.93	0.84

Note. For behavioral criteria, $N = 392$ and $R^2 \geq 0.03$ are statistically significant at $p < 0.05$. For well-being criteria, $N = 185$ and $R^2 \geq 0.06$ are statistically significant at $p < 0.05$. For peer-reported criteria, $N = 230$ and $R^2 \geq 0.05$ are statistically significant at $p < 0.05$. Proportion of BFI-2 domain scale validity retained equals the ratio of mean criterion variance explained by a short form compared to the full BFI-2.

had less overall predictive power than the full measure. Second, the BFI-2-S should not be used to examine facet-level associations in smaller samples. The well-being and peer-reported criteria were each assessed in samples of approximately 200 participants, and in these samples the BFI-2-S showed substantial discrepancies from the full BFI-2's pattern of facet-level associations³. Finally, the present results suggest that the BFI-2-XS should be used to assess personality only at the domain level. Across the full set of external criteria, there were substantial discrepancies between the facet-level associations observed for the full BFI-2 vs. the BFI-2-XS.

3.2.5. Facet-level measurement properties of the BFI-2-S

Because the correlation and regression analyses just described suggest that the BFI-2-S may be useful for assessing facet traits in reasonably large samples, our final set of analyses examined this short form's facet-level measurement properties. Table 6 presents part-whole correlations, alpha reliabilities, and self-peer agreement correlations for the BFI-2 and BFI-2-S facet scales. The leftmost part of this table shows that, in each validation sample, part-whole correlations for the BFI-2-S averaged 0.91 (total range = 0.86–0.95), indicating that this short form captures approximately 83% of the total variance in the full BFI-2 facet scales.

³ This conclusion was further supported by analyses of the BFI-2 and BFI-2-S facet scales' associations with the self-reported behavioral criteria in the subsamples of participants who were also assessed on the well-being ($N = 177$) and peer ($N = 225$) criteria. Associations between the facet scales and the behavioral criteria replicated somewhat less clearly between the full BFI-2 and the BFI-2-S in these smaller subsamples than in the full sample of 392 participants who completed the self-reported behavioral measure.

The middle part of Table 6 shows that, in each sample, alpha reliabilities of the full BFI-2 facet scales averaged 0.74 or 0.75 (total range = 0.59–0.84), whereas alphas of the BFI-2-S facets averaged 0.60 or 0.61 (total range = 0.39–0.79). As with the BFI-2-XS domain scales, the relatively low internal consistency of the BFI-2-S facet scales reflects their extreme brevity (2 items per scale) and our decision to maintain content validity by prioritizing breadth over internal consistency (cf. Gosling et al., 2003; Smith et al., 2000; Stanton et al., 2002).

Finally, the rightmost part of Table 6 shows that, in the university sample, self-peer agreement correlations averaged 0.49 for the full BFI-2 facet scales (range = 0.30–0.72), and 0.45 for the BFI-2-S facets (range = 0.23–0.69). These results indicate that the BFI-2-S retains approximately 85% of the full BFI-2 facet scales' self-peer agreement⁴.

4. General discussion

The present research pursued three main goals. First, we used a joint rational-empirical approach to develop two short forms of the Big Five Inventory–2: the 30-item BFI-2-S and the 15-item BFI-2-XS. Second, we examined how well these short forms retain the reliability and validity of the full BFI-2. At the level of the Big Five domains, our analyses of multiple indicators converge in showing that the BFI-2-S retains about 90%, and the BFI-2-XS about 80%, of the BFI-2 domain scales' reliability, self-peer agreement, and external validity (see Tables 1 and 5). Finally, we examined whether the BFI-2 short forms are appropriate for hierarchically assessing more-specific facet traits within the broad Big Five domains. Here our results suggest that the BFI-2-S may be useful for examining facet-level traits in reasonably large samples (approximately 400 or more observations), in which case we estimate that they retain approximately 85% of the full BFI-2 facet scales' reliability and validity (see Table 6). In contrast, our results indicate that the BFI-2-XS is so brief that it should only be used to assess personality at the level of the Big Five domains, not at the facet level.

4.1. When (and when not) to use the BFI-2 short forms

When designing a study and deciding whether to administer one of the BFI-2 short forms instead of the full measure, researchers should consider a few key points. First, the short forms do offer some savings in assessment time over the full BFI-2. Based on experiences with the original BFI and the BFI-2, we estimate that most participants can complete the full BFI-2 in four to ten minutes, the BFI-2-S in three to five minutes, and the BFI-2-XS in two or three minutes. These modest time savings may be important for certain studies in which minimizing assessment time and respondent fatigue are vital concerns. Some examples include large-scale surveys designed to efficiently assess many variables beyond personality traits, within-subjects designs that require each participant to complete the same personality measure multiple times, and laboratory studies that must reserve considerable time for experimental manipulations and behavioral observations. In these highly constrained contexts, administering the full BFI-2 may not be feasible, but administering one of the BFI-2 short forms would clearly be better than not measuring personality at all.

⁴ Our recommendation against using the BFI-2-XS to assess facet traits was further supported by analyses of this short form's facet-level measurement properties. For the single BFI-2-XS items, part-whole correlations averaged only .79 in each validation sample, and self-peer agreement correlations averaged only .38 in the university validation sample. These results indicate a substantial loss of reliability and self-peer agreement from the BFI-2 and BFI-2-S facet scales to the single BFI-2-XS items.

Table 6
Summary and comparison of measurement properties for the BFI-2 and BFI-2-S facet scales (Study 2).

Facet scale	Part-whole correlations BFI-2-S		Alpha reliabilities				Self-peer correlations	
			BFI-2		BFI-2-S		BFI-2	BFI-2-S
	Inter.	Univ.	Inter.	Univ.	Inter.	Univ.	Univ.	Univ.
<i>Extraversion</i>								
Sociability	0.95	0.95	0.83	0.84	0.70	0.71	0.72	0.69
Assertiveness	0.91	0.92	0.76	0.75	0.72	0.67	0.57	0.50
Energy Level	0.87	0.89	0.70	0.69	0.60	0.49	0.49	0.50
<i>Agreeableness</i>								
Compassion	0.87	0.86	0.59	0.65	0.48	0.55	0.38	0.41
Respectfulness	0.91	0.89	0.70	0.69	0.48	0.48	0.37	0.37
Trust	0.90	0.90	0.68	0.71	0.53	0.60	0.48	0.44
<i>Conscientiousness</i>								
Organization	0.94	0.94	0.83	0.82	0.79	0.75	0.56	0.52
Productiveness	0.92	0.88	0.77	0.70	0.58	0.45	0.51	0.44
Responsibility	0.90	0.89	0.71	0.62	0.47	0.39	0.32	0.23
<i>Negative Emotionality</i>								
Anxiety	0.92	0.91	0.77	0.77	0.65	0.67	0.56	0.48
Depression	0.94	0.94	0.81	0.80	0.67	0.66	0.50	0.50
Emotional Volatility	0.94	0.94	0.83	0.83	0.75	0.76	0.56	0.50
<i>Open-Mindedness</i>								
Intellectual Curiosity	0.89	0.91	0.66	0.73	0.42	0.57	0.52	0.45
Aesthetic Sensitivity	0.90	0.93	0.73	0.82	0.54	0.69	0.48	0.38
Creative Imagination	0.92	0.91	0.75	0.77	0.64	0.65	0.30	0.31
Mean coefficient	0.91	0.91	0.74	0.75	0.60	0.61	0.49	0.45
Proportion of BFI-2 facet scale variance retained	0.83	0.83	–	–	0.81	0.81	–	0.85

Note. Inter. = Internet validation sample ($N = 2000$). Univ. = University validation sample ($N = 423$; for self-peer correlations, $N = 230$ and all are statistically significant at $p < 0.001$). Proportion of BFI-2 facet scale variance retained equals the mean squared part-whole correlation, the ratio of the mean alpha reliability for the BFI-2-S compared to the full BFI-2, or the ratio of the mean squared self-peer correlation for the BFI-2-S compared to the full BFI-2.

However, it is important to recognize that the short forms' gains in efficiency come at a cost to reliability and validity. We have attempted to reduce such costs by drawing on a combination of rational and empirical criteria to carefully select items for the short forms. Even so, at the level of the Big Five domains the present results indicate that the BFI-2-S provides approximately 10%, and the BFI-2-XS approximately 20%, less reliability and validity than the full BFI-2 domain scales. We should therefore expect to observe systematically smaller associations between personality traits and external criteria when using a short form rather than the full BFI-2 (John & Soto, 2007). An important but easily overlooked implication of such shrinkage is that studies administering abbreviated measures will need to recruit larger samples of participants to maintain adequate statistical power. For example, imagine that we expect a zero-order correlation of 0.20 between a BFI-2 domain scale and an external criterion (a typical effect size in psychological research; Richard, Bond, & Stokes-Zoota, 2003). For 80% power to detect this effect, we would need to recruit approximately 190 participants if administering the full BFI-2, 220 participants if administering the BFI-2-S, or 240 participants if administering the BFI-2-XS.

The decreased reliability of the BFI-2 short forms also has important implications for testing more complex effects. One common example is research assessing whether a newly proposed psychological construct can provide incremental validity beyond established constructs (e.g., the Big Five) treated as control variables. Assessing the established control variables with a high degree of reliability helps prevent spurious evidence of incremental validity (Westfall & Yarkoni, 2016). Similar reliability concerns apply to mediation analyses, in which regressions or partial correlations are used to test whether the effect of a predictor variable on an outcome can be accounted for by a mediator (Baron & Kenny, 1986). We therefore recommend that researchers administer the full BFI-2 when evaluating the incremental validity of constructs

beyond the Big Five domains and facets, or when testing mediation relationships.

Finally, researchers should consider whether they wish to examine personality traits hierarchically, by distinguishing between more-specific facet traits within each broad Big Five domain. Such hierarchical assessment combines the benefits of high bandwidth at the domain level with high fidelity at the facet level, and can therefore substantially enhance a measure's capacity to accurately predict a wide range of external criteria (Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Paunonen & Ashton, 2001; Soto & John, in press). All three forms of the BFI-2 are appropriate for domain-level personality assessment, but only the full BFI-2 and the BFI-2-S (in reasonably large samples) are also appropriate for examining facet-level traits. The BFI-2-XS' lack of hierarchical assessment entails a further cost to validity: the BFI-2-XS domain scales provide only about two-thirds as much predictive power as the BFI-2 facet scales. Thus, for example, if the BFI-2 facets explain 50% of the variance in a particular criterion, then we would only expect the BFI-2-XS domains to explain about 33% of the criterion variance.

We therefore advise researchers to carefully weigh the relative values of efficiency, reliability, validity, bandwidth, fidelity, and statistical power when deciding which form of the BFI-2 to administer in a particular study. For example, would saving four minutes of assessment time by administering the BFI-2-XS instead of the full BFI-2 be worth increasing a study's planned sample size by 25% to maintain statistical power, decreasing the size of the observed effects by 10%, and losing the capacity to examine facet-level traits? For many studies, the BFI-2 short forms' modest time savings will clearly be outweighed by the full measure's superior measurement properties; we therefore strongly recommend administering the full BFI-2 in most research contexts. However, in studies where assessment time and respondent fatigue are core concerns that cannot feasibly be mitigated by other design

decisions, the BFI-2 short forms' small gains in efficiency may be crucially important; in such cases, the short forms are useful alternatives to the full measure.

4.2. Limitations and future directions

The present research had a number of important strengths, including its joint rational-empirical approach to scale construction, its use of multiple independent samples to directly replicate key results, and its assessment of both self-reported and peer-reported validity criteria. However, it also had some limitations that highlight important directions for future research. First, the retest reliabilities for the BFI-2-S and BFI-2-XS observed in Study 1 may be positively biased because they were computed in the same samples that we used to select items for these short forms. Second, we administered the short forms embedded within the full BFI-2, rather than administering each form separately (cf. Rammstedt & John, 2007; Smith et al., 2000). Additional research is therefore needed to further examine the short-term retest reliability and long-term stability of the BFI-2-S and BFI-2-XS, and to test whether the short forms' measurement properties are affected by separate vs. embedded administration.

Third, our current recommendation that the BFI-2-S may be useful for examining facet-level traits in reasonably large samples (approximately 400 or more observations), and that the BFI-2-XS is not appropriate for examining facet traits, should be considered provisional. Our recommendation is based on analyses of 20 criterion variables assessed in three partially overlapping subsamples. Additional research using larger samples and broader sets of criteria could provide more definitive conclusions. For example, future studies can test whether the BFI-2-XS may be useful for examining facet traits in very large samples. The gains in statistical power and precision that come from samples numbering in the thousands rather than hundreds of observations may adequately compensate for the limited reliability entailed by using only a single item to assess each facet (cf. Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002).

Finally, additional research is needed to compare the BFI-2 short forms with other brief Big Five measures, including the Ten

Item Personality Inventory (TIPI; Gosling et al., 2003) and the Mini-IPIP (Donnellan, Oswald, Baird, & Lucas, 2006). These brief measures differ in terms of their length (ranging from 10 to 30 items), item format (adjective pairs for the TIPI vs. short phrases for the Mini-IPIP and BFI-2), and scale development strategy (prioritizing simple factor structure for the Mini-IPIP vs. content breadth for the TIPI and BFI-2). Comparative research can therefore examine how these differences affect the measures' reliability and validity.

5. Conclusion

The BFI-2 is a 60-item questionnaire that hierarchically assesses the Big Five personality domains and 15 more-specific facet traits. The present research developed and validated a 30-item short form (the BFI-2-S) and a 15-item extra-short form (the BFI-2-XS) of the BFI-2. These abbreviated forms offer savings in assessment time over the full BFI-2. They also retain most of the full measure's reliability and validity, especially at the domain level. The BFI-2-S and BFI-2-XS should therefore prove useful for assessing personality traits in research contexts where, due to pressing concerns about assessment time or respondent fatigue, administering the full BFI-2 would not be feasible. For most studies, however, we recommend administering the full measure due to its greater reliability and validity.

Acknowledgment

This research was supported by faculty research grants from Colby College to Christopher J. Soto, and from the University of California, Berkeley to Oliver P. John. The authors thank Seth Butler, Emma Heilbronner, Sara Heilbronner, Caroline Minott, Natalia Van Doren, and Carylanne Wolfington for providing access to some of the college and university sample data. For downloadable versions of the BFI-2 and its short forms, visit the Colby Personality Lab website (<http://www.colby.edu/psych/personality-lab/>). To initiate a translation of the BFI-2 or its short forms, please contact the authors.

Appendix A. The BFI-2-S and BFI-2-XS

A.1. The BFI-2 short forms and scoring information

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1	2	3	4	5
Disagree strongly	Disagree a little	Neutral; no opinion	Agree a little	Agree strongly

I am someone who...

1. <input type="checkbox"/> Tends to be quiet.	16. <input type="checkbox"/> Is outgoing, sociable.
2. <input type="checkbox"/> Is compassionate, has a soft heart.	17. <input type="checkbox"/> Can be cold and uncaring.
3. <input type="checkbox"/> Tends to be disorganized.	18. <input type="checkbox"/> Keeps things neat and tidy.
4. <input type="checkbox"/> Worries a lot.	19. <input type="checkbox"/> Is relaxed, handles stress well.
5. <input type="checkbox"/> Is fascinated by art, music, or literature.	20. <input type="checkbox"/> Has few artistic interests.
6. <input type="checkbox"/> Is dominant, acts as a leader.	21. <input type="checkbox"/> Prefers to have others take charge.
7. <input type="checkbox"/> Is sometimes rude to others.	22. <input type="checkbox"/> Is respectful, treats others with respect.

(continued on next page)

Appendix A. (continued)

8.	___ Has difficulty getting started on tasks.	23.	___ Is persistent, works until the task is finished.
9.	___ Tends to feel depressed, blue.	24.	___ Feels secure, comfortable with self.
10.	___ Has little interest in abstract ideas.	25.	___ Is complex, a deep thinker.
11.	___ Is full of energy.	26.	___ Is less active than other people.
12.	___ Assumes the best about people.	27.	___ Tends to find fault with others.
13.	___ Is reliable, can always be counted on.	28.	___ Can be somewhat careless.
14.	___ Is emotionally stable, not easily upset.	29.	___ Is temperamental, gets emotional easily.
15.	___ Is original, comes up with new ideas.	30.	___ Has little creativity.

Please check: Did you write a number in front of each statement?

Note. BFI-2 items copyright 2015 by Oliver P. John and Christopher J. Soto. Reprinted with permission.

A.2. Scoring the BFI-2-S domain and facet scales

The BFI-2-S includes items 1–30 shown on the above administration form. Item numbers for scoring the BFI-2-S domain and facet scales are listed below; false-keyed items are denoted by “R.” Due to the limited reliability of the two-item facet scales, we only recommend using them in samples with approximately 400 or more observations. For downloadable versions and more information about the BFI-2, visit the Colby Personality Lab website (<http://www.colby.edu/psych/personality-lab/>).

Extraversion: 1R, 6, 11, 16, 21R, 26R

Sociability: 1R, 16
Assertiveness: 6, 21R
Energy Level: 11, 26R

Agreeableness: 2, 7R, 12, 17R, 22, 27R

Compassion: 2, 17R
Respectfulness: 7R, 22
Trust: 12, 27R

Conscientiousness: 3R, 8R, 13, 18, 23, 28R

Organization: 3R, 18
Productiveness: 8R, 23
Responsibility: 13, 28R

Negative Emotionality: 4, 9, 14R, 19R, 24R, 29

Anxiety: 4, 19R
Depression: 9, 24R
Emotional Volatility: 14R, 29

Open-Mindedness: 5, 10R, 15, 20R, 25, 30R

Aesthetic Sensitivity: 5, 20R
Intellectual Curiosity: 10R, 25
Creative Imagination: 15, 30R

A.3. Administering and scoring the BFI-2-XS domain scales

The BFI-2-XS is the first half of the BFI-2-S. Therefore, the BFI-2-XS includes items 1–15 shown on the above BFI-2-S administration form, and the BFI-2-XS domain scales are scored from the first three items listed for each domain scale on the above BFI-2-S scoring key. Due to the limited reliability of single items, we do not recommend using the BFI-2-XS to examine facet-level traits.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jrp.2017.02.004>.

References

- Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality, 53*, 1–4.
- Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality, 29*, 432–442.
- Bardi, A., & Schwartz, S. H. (2003). Values and behavior: Strength and structure of relations. *Personality and Social Psychology Bulletin, 29*, 1207–1220.
- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology, 69*, 49–56.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality, 57*, 119–130.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880–896.
- 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*, 192–203.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.
- Goldberg, L. R., & Kilkowski, J. M. (1985). The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology, 48*, 82–98.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504–528.
- Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin, 55*, 243–252.
- John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). Is there a basic level of personality description? *Journal of Personality and Social Psychology, 60*, 348–361.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York, NY: Guilford.
- John, O. P., & Soto, C. J. (2007). The importance of being valid. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461–494). New York, NY: Guilford.
- McCrae, R. R., & Costa, P. T. (2010). *NEO inventories professional manual*. Lutz, FL: Psychological Assessment Resources.
- McCrae, R. R., & Costa, P. T. (2008). The Five-Factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 159–181). New York, NY: Guilford.
- Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*, 524–539.
- Pozzobon, J. A., & Ashton, M. C. (2009). Personality and values as predictors of self- and peer-reported behavior. *Journal of Individual Differences, 30*, 122–129.
- Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment, 25*, 1137–1145.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*, 203–212.
- Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*, 331–363.
- Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*, 103–139.
- Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17*, 423–434.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*, 1069–1081.

- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 102–111.
- Soto, C. J., & John, O. P. (in press). 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*.
- 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 ten to twenty. *Journal of Personality and Social Psychology*, *94*, 718–737.
- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, *98*, 520–534.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, *55*, 167–194.
- Summerfield, M., Freidin, S., Hahn, M., Li, N., Macalalad, N., Mundy, L., ... Wooden, M. (2015). *HILDA user manual: Release 14*. Melbourne, Australia: Melbourne Institute of Applied Economic and Social Research, University of Melbourne.
- Taylor, M. F., Brice, J., Buck, N., & Prentice-Lane, E. (2010). *British household panel survey user manual: Volume A. Introduction, technical report and appendices*. Colchester, England: University of Essex.
- Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): Scope, evolution and enhancements. *Schmollers Jahrbuch*, *127*, 139–169.
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS One*, *11*, e0152719.
- Wood, D., & Roberts, B. W. (2006). Cross-sectional and longitudinal tests of the Personality and Role Identity Structural Model (PRISM). *Journal of Personality*, *74*, 779–809.