Journal of Personality and Social Psychology

The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model With 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power

Christopher J. Soto and Oliver P. John

Online First Publication, April 7, 2016. http://dx.doi.org/10.1037/pspp0000096

CITATION

Soto, C. J., & John, O. P. (2016, April 7). 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*. Advance online publication. http://dx.doi.org/10.1037/pspp0000096

Journal of Personality and Social Psychology
© 2016 American Psychological Association
0022-3514/16/$12.00 http://dx.doi.org/10.1037/pspp0000096

The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model With 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power

Christopher J. Soto
Colby College

Oliver P. John
University of California, Berkeley

Three studies were conducted to develop and validate the Big Five Inventory-2 (BFI-2), a major revision of the Big Five Inventory (BFI). Study 1 specified a hierarchical model of personality structure with 15 facet traits nested within the Big Five domains, and developed a preliminary item pool to measure this structure. Study 2 used conceptual and empirical criteria to construct the BFI-2 domain and facet scales from the preliminary item pool. Study 3 used data from 2 validation samples to evaluate the BFI-2's measurement properties and substantive relations with self-reported and peer-reported criteria. The results of these studies indicate that the BFI-2 is a reliable and valid personality measure, and an important advance over the original BFI. Specifically, the BFI-2 introduces a robust hierarchical structure, controls for individual differences in acquiescent responding, and provides greater bandwidth, fidelity, and predictive power than the original BFI, while still retaining the original measure's conceptual focus, brevity, and ease of understanding. The BFI-2 therefore offers valuable new opportunities for research examining the structure, assessment, development, and life outcomes of personality traits.

Keywords: 5-factor model, Big Five, facets, personality measurement, personality structure

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Many important individual differences in people's patterns of thinking, feeling, and behaving can be summarized in terms of the Big Five personality domains, which we label Extraversion, Agreeableness, Conscientiousness, Negative Emotionality (alternatively labeled Neuroticism vs. Emotional Stability), and Open-Mindedness (alternatively labeled Openness to Experience, Intellect, or Imagination; Goldberg, 1993b; John, Naumann, & Soto, 2008; McCrae & Costa, 2008). Over the past quarter century, scientific consensus regarding the structure and basic definitions of the Big Five has generated an explosion of research documenting their causes, correlates, and consequences. A considerable portion of this research has measured personality traits using the Big Five Inventory (BFI), which assesses the prototypical features of each Big Five domain using 44 short and easy-to-understand phrases (John, Donahue, & Kentle, 1991; see John & Srivastava, 1999; John et al., 2008). The BFI has been used in hundreds of studies to date, and has demonstrated considerable reliability, validity, and utility. However, the 25 years since the BFI's original development have also yielded important advances in our understanding of both personality structure and psychological assessment. Thus, the present research was conducted to integrate these advances into the BFI, while simultaneously addressing a number of substantive and measurement issues. This research program culminates with the development and initial validation of the Big Five Inventory-2 (BFI-2).

The Big Five Inventory: History and Key Features

The original BFI was developed with three key goals in mind. First, it was designed to focus on the prototypical components of each Big Five domain. This focus was achieved through a combined rational-empirical approach to scale construction. Specifically, a panel of expert judges reviewed the 300 items of the Adjective Check List (ACL; Gough & Heilbrun, 1983) and identified those that were conceptually relevant to one of the Big Five. Analyses of observer-reports were then conducted to test and refine the selected adjectives' Big Five structure. This two-stage process yielded a set of approximately 100 trait-descriptive adjectives that were conceptually and empirically central to the Big Five, and these adjectives provided a lexical foundation from which to develop the BFI (John, 1989, 1990).

A second key goal for the BFI was clarity. Because trait-descriptive adjectives are understood and rated more consistently when they are accompanied by elaborations or definitions (Goldberg & Kilkowski, 1985), the prototypical Big Five adjectives identified by John (1989, 1990) were elaborated into short phrases for the BFI. Specifically, most BFI items are structured in one of

Christopher J. Soto, Department of Psychology, Colby College; Oliver P. John, Department of Psychology and Institute of Personality and Social Research, University of California, Berkeley.

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. For a downloadable version of the BFI-2 and more information, visit the Colby Personality Lab website (http://www.colby.edu/psych/personality-lab/). To initiate a translation, short form, or other adaptation of the BFI-2, please contact the authors. The authors thank Juliana Pham for her assistance with this research, and Wiebke Bleidorn for her helpful comments on a draft of this article.

Correspondence concerning this article should be addressed to Christopher J. Soto, Department of Psychology, Colby College, 5550 Mayflower Hill, Waterville, ME 04901. E-mail: christopher.soto@colby.edu

three ways: (a) adjective, synonym (e.g., "Is outgoing, sociable"), (b) adjective, definition (e.g., "Is relaxed, handles stress well"), or (c) adjective in context (e.g., "Is a reliable worker"). These short phrases retain the simplicity and brevity of adjectival items, while addressing the limitation that individual adjectives often have ambiguous or multiple meanings.

A final key goal was efficiency. Each BFI scale is long enough to be reliable and provide reasonably broad coverage of its Big Five domain, but still short enough to conserve research time and prevent respondent fatigue. At 44 total items, the BFI can be completed in 5 to 10 min. At the time of its development, the BFI was considerably shorter than most available broadband personality inventories, which typically included hundreds of items.

We sought to retain these three key strengths—focus, clarity, and brevity—while developing the BFI-2. Specifically, we wanted the BFI-2 to capture the prototypical features of each Big Five domain, to use easy-to-understand phrases, and to be short enough for research participants to complete in 10 min or less. However, we also aimed to address some limitations of the original BFI by integrating important advances from research on personality structure and psychological assessment.

Personality Structure and Assessment: Bandwidth, Fidelity, and Hierarchy

Personality traits can be defined with different degrees of conceptual breadth. A broadly defined trait (e.g., Conscientiousness) has the advantage of high bandwidth: it efficiently summarizes a large amount of behavioral information, and can predict a variety of relevant criteria (John et al., 2008; Ozer & Benet-Martínez, 2006). Conversely, a narrowly defined trait (e.g., Organization) has the advantage of high fidelity: it provides a more precise description of behavior, and can predict closely matched criteria with greater accuracy (Ashton, Jackson, Paunonen, Helmes, & Rothstein, 1995; Paunonen & Ashton, 2001). The fact that traits with different degrees of breadth have different advantages and disadvantages is known as the bandwidth-fidelity tradeoff (Cronbach & Gleser, 1957; John, Hampson, & Goldberg, 1991).

The bandwidth-fidelity tradeoff has two key implications for conceptualization and measurement of the Big Five. First, one possible solution to this tradeoff is hierarchical assessment. Each broad Big Five domain can be conceptualized as subsuming several more-specific "facet" traits, and a single instrument can simultaneously assess personality at both the domain and facet levels (Costa & McCrae, 1995). Domain-level scales provide the breadth and efficiency of high bandwidth, whereas facet-level subscales provide the descriptive and predictive precision of high fidelity. A growing body of research has adopted this hierarchical approach. For example, Costa and McCrae (1995) developed a set of 30 facets (6 per Big Five domain) from a conceptual review of the personality literature followed by additional psychometric research. Hofstee, De Raad, and Goldberg (1992) used a circumplex approach to define an even larger set of 45 facets (9 per domain) within their Abridged Big Five Circumplex (AB5C) model. Saucier and Ostendorf (1999) empirically examined the hierarchical structure of trait-descriptive adjectives in English and German, and identified 18 facet-level subcomponents (3 or 4 within each domain) that replicated between the two languages. Each of these early projects identified a somewhat different set of Big Five facets, but more recent work has begun to integrate their results. For example, one influential project (DeYoung, Quilty, & Peterson, 2007) used factor analysis to derive two major and maximally independent "aspects" of each Big Five domain that represent commonalities among the 75 total facets proposed by Costa and McCrae (1995) and Hofstee et al. (1992). Similarly, a broad review of the Big Five literature (John et al., 2008) highlighted overlaps between the facets conceptually defined by Costa and McCrae (1995), the lexical subcomponents empirically identified by Saucier and Ostendorf (1999), and the AB5C facets best represented on the classic California Psychological Inventory (Gough, 1957; see Soto & John, 2009b). These integrative papers suggest that, within each Big Five domain, a large amount of more-specific personality information can be captured by two to four facet traits that consistently replicate across alternative hierarchical models.

The original BFI was developed before all of these projects, and was therefore not intended as a hierarchical measure. However, it is now clear that facet-level traits capture meaningful personality information, that many facets show distinctive developmental trends, and that individual facets relate uniquely with important behaviors and life outcomes (e.g., Ashton et al., 1995; Costa & McCrae, 1995; Hirsh, DeYoung, Xiaowen Xu, & Peterson, 2010; Paunonen & Ashton, 2001; Soto, John, Gosling, & Potter, 2011; Terracciano, McCrae, Brant, & Costa, 2005). In a previous project, we therefore developed post hoc facet scales for the original BFI (Soto & John, 2009a). These post hoc facet scales have proven useful in a number of studies (e.g., Markowitz, Goldberg, Ashton, & Lee, 2012; Soto et al., 2011), but they also have important limitations. Most notably, each BFI domain scale has only 8 to 10 items, and these items were not developed with facet-level distinctions in mind. Thus, the original BFI includes only limited content relevant to many Big Five facets. In constructing the BFI-2, we therefore aimed to develop a robust hierarchical structure with reliable domain and facet scales that can jointly offer the advantages of high bandwidth and high fidelity.

A second important, but less frequently noted, implication of bandwidth and fidelity is that the Big Five domains themselves can be conceptualized and assessed more broadly or more narrowly. For example, a narrowly defined version of Conscientiousness could focus on a single, central facet, such as Organization. A moderately broad version of Conscientiousness could add a few complementary facets such as Productiveness and Responsibility, whereas a very broad version could add even more peripheral facets such as conventionality and virtue (DeYoung et al., 2007; Hofstee et al., 1992; MacCann, Duckworth, & Roberts, 2009; Roberts, Bogg, Walton, Chernyshenko, & Stark, 2004; Roberts, Chernyshenko, Stark, & Goldberg, 2005). Similarly, any individual Big Five facet can be defined as including a relatively broad or narrow range of lower-level "nuance" traits (McCrae, 2015). For example, a narrow version of Organization could focus exclusively on preference for Order and structure, while a broader version of this facet could include additional nuances such as neatness and attention to detail.

The choice to define and assess a relatively broad or narrow version of a particular trait will affect the resulting measure's psychometric properties and substantive relations with external criteria. For example, a narrow measure of a Big Five domain will tend to have greater overlap between its items (because these items will all assess closely related patterns of behavior), and therefore

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

stronger interitem correlations and greater internal consistency reliability (John & Soto, 2007). Moreover, an inventory that narrowly assesses each Big Five domain will tend to have a clear and simple five-factor structure, because greater internal consistency within each domain will enhance primary factor loadings, whereas the exclusion of peripheral facets—which often combine content from multiple Big Five domains (Hofstee et al., 1992)—will suppress secondary loadings. Conversely, an inventory that broadly assesses each Big Five domain will tend to have lower internal consistency and a more complex factor structure due to its inclusion of peripheral content, but this same peripheral content will also allow it to accurately predict a wider range of criteria. These same considerations also apply at the facet level. Scales measuring narrowly defined facets will benefit from high internal consistency and a clear facet-level factor structure, but overly narrow facets will have limited predictive relevance (John & Soto, 2007; McCrae, 2015).

The original BFI was developed from trait-descriptive adjectives that expert judges identified as clearly related to one—and only one—Big Five domain. This strict inclusion criterion resulted in tightly focused domain scales with correspondingly strong internal consistency and a clear Big Five structure (John et al., 2008). However, in developing the BFI-2 we aimed to somewhat broaden the five domains' conceptualization and measurement. We expected that this greater breadth would provide a more comprehensive representation of each domain's content and enhance their substantive relations with external criteria. At the facet level, we pursued the opposite goal. Because the original BFI was developed without facet-level distinctions in mind, it represents some facets more extensively than others (John et al., 2008). For example, the BFI Conscientiousness scale includes more items related to Productiveness than to Organization. Moreover, some items include content relevant to multiple facets (e.g., "Does a thorough job" is relevant to both Productiveness and Organization). When constructing post hoc facet scales for the original BFI, this led some facets to be overly broad and poorly differentiated (Soto & John, 2009a). In developing the BFI-2, we aimed to draw sharper distinctions between facets, to increase differentiation between the facet scales and clarify the measure's hierarchical structure.

Acquiescent Response Style: Problems and Solutions

Acquiescent response style is the tendency of an individual to consistently agree (yea-saying) or consistently disagree (nay-saying) with questionnaire items, regardless of their content (Jackson & Messick, 1958). The presence of individual differences in acquiescence can distort the reliability, validity, and structure of a psychological measure at both the item and scale levels. At the item level, acquiescence variance tends to positively bias interitem correlations: positive correlations (e.g., between two true-keyed items on the same scale) tend to become stronger, whereas negative correlations (e.g., between a true-keyed item and a false-keyed item) tend to become weaker. These biased correlations, in turn, affect the item-level factor structure; specifically, they tend to pull true-keyed and false-keyed items toward separate factors, and can even lead to the emergence of an additional method factor representing acquiescence (McCrae, Herbst, & Costa, 2001; Ten Berge, 1999). Such item-level acquiescence effects have been documented for the BFI and many other psychological measures (e.g., John et al., 2008; Ten Berge, 1999). Acquiescence effects are particularly pronounced in samples of children and adolescents, as well as adults with low levels of educational attainment, because individual differences in acquiescence tend to be greater in these groups (Rammstedt & Farmer, 2013; Soto, John, Gosling, & Potter, 2008).

At the scale level, individual differences in acquiescence can affect internal consistency reliability (Soto et al., 2008). Specifically, acquiescence variance tends to inflate the internal consistency of scales with a substantial imbalance of true-keyed and false-keyed items by positively biasing correlations between pairs of like-keyed items. Acquiescence can also bias correlations between pairs of imbalanced scales. An interscale correlation will be positively biased if both scales have a majority of true-keyed items (or of false-keyed items), and negatively biased if one scale has mostly true-keyed items while the other has mostly false-keyed items. The original BFI has relatively balanced scales for four of the Big Five domains: Extraversion, Agreeableness, Conscientiousness, and Negative Emotionality. Each of these scales has five true-keyed items and three or four false-keyed items. However, reflecting the fact that many more adjectives in natural language describe high levels than low levels of intellectual and creative interests (Hofstee et al., 1992; John, 1989, 1990), the BFI Open-Mindedness scale includes eight true-keyed items and only two false-keyed items. Moreover, some of the post hoc BFI facet scales include only true-keyed or only false-keyed items (Soto & John, 2009a). These imbalanced scales are particularly susceptible to the influence of acquiescence.

To address this issue, we aimed to construct content-balanced scales for the BFI-2 at both the domain and facet levels. Balancing the number of true-keyed and false-keyed items on each scale would clearly distinguish meaningful personality information from acquiescence. Specifically, balanced keying would automatically control for acquiescence at the scale level, because yea-saying (or nay-saying) in response to a scale's true-keyed items would be canceled out by responding similarly to its equal number of false-keyed items. Content-balanced scales would also allow researchers to easily control for acquiescence at the item level by centering each individual's set of item responses around their within-person mean (i.e., their mean response to the full item set, without reversing the false-keyed items; see Soto et al., 2008).

Adopting More Easily Understood Labels for N and O

While pursuing these goals for the BFI-2, we decided to make one additional change from the original BFI. Specifically, the BFI-2 adopts new labels for two of the Big Five domains. Researchers have long acknowledged the difficulty of choosing accurate and easily understood labels for the Big Five, because each of these broad personality domains "is not so much one thing as a collection of many things that have something in common" (Saucier & Goldberg, 2003, p. 14). This challenge has proven particularly difficult for the domains traditionally labeled Neuroticism and Openness to Experience.

The label Neuroticism is rooted in the term neurosis, which first (in 18th and 19th century medicine) connoted a general disorder of the nervous system, and later (in early 20th century psychiatry) a type of mental illness characterized by psychological distress (Widiger, 2009). Since the mid-20th century, personality research-

ers have used the term Neuroticism to describe an individual's general tendency to experience negative emotions such as Anxiety and sadness (Eysenck, 1967; John et al., 2008). However, in everyday language the adjective "neurotic" retains its clinical connotation (Merriam-Webster, 2005). The BFI-2 therefore adopts the label Negative Emotionality (Clark & Watson, 2008) to highlight this domain's focus on negative emotional experiences while more clearly distinguishing it from psychiatric illness.

The term Openness to Experience was coined by Rogers (1954; see also McCrae & Costa, 2008) to describe the extent to which an individual prefers to have a wide versus narrow range of perceptual, cognitive, and affective experiences. Compared with alternative labels such as Intellect, Imagination, or Culture, Openness more fully conveys the broad scope of this personality domain and more clearly distinguishes it from intelligence (McCrae, 1994). However, nonpsychologists often misinterpret it as connoting openness to social experiences (i.e., Extraversion) rather than mental experiences (Sneed, McCrae, & Funder, 1998). For the BFI-2, we therefore adopt the label Open-Mindedness, which retains Openness' sense of breadth but clarifies this domain's focus on an individual's mental rather than social life.

Overview of the Present Research

In sum, we developed the BFI-2 with four major goals in mind. First, we aimed to establish a robust hierarchical structure, with multiple facet traits nested within each Big Five domain. Second, we aimed to balance bandwidth and fidelity at both the domain and facet levels, in order to enhance descriptive and predictive power. Third, we aimed to minimize the influence of acquiescence through the construction of content-balanced scales. Finally, we aimed to retain the focus, clarity, and brevity of the original BFI. We pursued these goals through a program of three studies. Study 1 conceptually defined a hierarchical structure for the BFI-2 and developed a preliminary item pool to measure this structure. Study 2 refined this preliminary pool into the final BFI-2 domain and facet scales. Study 3 examined the basic measurement properties, multidimensional structure, and nomological network of the BFI-2. Collectively, these studies drew on data from community, student, and Internet samples, from a number of established personality measures, and from self-reported and peer-reported validity criteria.

Study 1

Study 1 was conducted to define the BFI-2's hierarchical structure—by identifying a set of conceptually and empirically prominent facet traits within each Big Five domain—and to develop a preliminary BFI-2 item pool that could be used to measure this hierarchical structure. Ideally, this item pool should be broad, deep, and balanced, with each domain and facet represented by several candidate items.

Method

Participants and procedure. Study 1 analyzed data from 1,137 members of the Eugene-Springfield Community Sample (ESCS), a sample of adults residing in Oregon (see Goldberg, 1999). The present sample ranged in age from 18 to 89 years old (M = 49.64, SD = 13.02), and included an approximately equal number of men and women (52.9% female, 46.5% male, 0.6% did not report gender). Reflecting the local demographics, almost all (97.7%) described themselves as White/Caucasian. Over several years, ESCS members have completed a number of psychological measures in exchange for compensation. The measures used in the present study are described below.

Measures.

Big Five Inventory. As described above, the original Big Five Inventory (BFI; John, Donahue, & Kentle, 1991; see John et al., 2008) was developed to measure the prototypical features of each Big Five domain. Its 44 items are short, descriptive phrases that respondents rate on a 5-point scale ranging from *disagree strongly* to *agree strongly*. A total of 641 ESCS members completed the BFI; in this sample, the domain scales' alpha reliabilities were .86 for Extraversion, .82 for Agreeableness, .83 for Conscientiousness, .85 for Negative Emotionality, and .84 for Open-Mindedness.

Trait-descriptive adjectives. Members of the ESCS rated themselves using a total of 885 trait-descriptive adjectives. These adjectives included 100 unipolar Big Five markers (Goldberg, 1992), the 40 Big Five Mini-Markers (Saucier, 1994), and a set of the 500 most familiar English trait-descriptive adjectives (Saucier, 1997), among others. The adjectives were administered in two lists: a 360-item list rated using a 9-point scale, and a 525-item list rated using a 7-point scale. Both rating scales ranged from *extremely inaccurate* to *extremely accurate* as a description of the respondent. In the present sample, 1,131 participants completed at least one of the two adjective lists, and most participants (685) completed both.

International Personality Item Pool. The International Personality Item Pool (IPIP; Goldberg, 1999; Goldberg et al., 2006) is a large, publicly available set of personality-relevant questionnaire items. Approximately 750 IPIP items were translated from Dutch to English (Hendriks, 1997), approximately 500 more (Goldberg, 1999) were developed to measure the Abridged Big Five Circumplex model (AB5C; Hofstee et al., 1992), and several additional, smaller item sets have been added to the IPIP over time (Goldberg et al., 2006). All of the IPIP items are short phrases rated using a 5-point scale ranging from *very inaccurate* to *very accurate* as a description of the respondent. The present study examined a total of 2,552 IPIP items administered to the ESCS in 12 lists. A total of 995 participants completed at least one of these lists, with 401 participants completing all 12 lists.

Results and Discussion

This study proceeded in three phases. First, we defined the BFI-2's hierarchical structure by identifying a set of conceptually and empirically prominent facet traits within each Big Five domain. Second, we identified potential item content relevant to measuring this hierarchical structure by analyzing trait-descriptive adjectives and phrases. Third, we used this content to construct a pool of candidate items for the BFI-2.

Phase 1: Selecting and defining facets. Our first step was to select and conceptually define a set of facet traits within each Big Five domain. Previous work has found that approximately two to four lower-level traits consistently replicate across alternative hierarchical models of the Big Five (John et al., 2008), and that this relatively small number of traits can capture a large amount of

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

specific personality information (DeYoung et al., 2007). We therefore defined three facets per BFI-2 domain. Our selection of facets was guided by previous research examining hierarchical personality structure and measurement (e.g., DeYoung et al., 2007; Goldberg, 1999; Hofstee et al., 1992; McCrae & Costa, 2008; Saucier & Ostendorf, 1999). Specifically, for each Big Five domain we began by selecting a factor-pure facet: a facet that previous research has identified as central to its own domain and independent from the other four domains (Hofstee et al., 1992). These factor-pure facets should empirically anchor the Big Five domains in multidimensional space. For each domain, we then selected two complementary facets that are prominent in the personality literature and represented in the original BFI's item content. These complementary facets should conceptually broaden each domain and provide continuity with the original BFI and previous research on personality structure. Table 1 lists the 15 BFI-2 facets alongside similar traits proposed by four influential hierarchical models of the Big Five (DeYoung et al., 2007; Hofstee et al., 1992; McCrae & Costa, 2008; Saucier & Ostendorf, 1999).

Previous research on personality structure has consistently identified Sociability (desire to socially approach and engage with others) as a facet central to Extraversion and orthogonal to the other Big Five domains (Goldberg, 1999; Hofstee et al., 1992; McCrae & Costa, 2008; Saucier & Ostendorf, 1999). Assertiveness (willingness to express personal opinions and goals in social situations) is a second prominent facet, and is empirically distinct from Sociability (DeYoung et al., 2007). Finally, Extraversion has been consistently linked with positive affect (especially positively aroused states such as Enthusiasm and excitement) and physical activity level, and these links cannot be fully explained by Sociability (Lucas, Le, & Dyrenforth, 2008; Watson & Clark, 1997). We therefore selected Sociability, Assertiveness, and Energy Level as the three BFI-2 Extraversion facets.

Prosocial emotion, cognition, and behavior are closely related components of Agreeableness that influence, but can also be somewhat distinguished from, each other (Graziano & Tobin, 2009). Lexical research suggests Compassion (active emotional concern for others' well-being) as a particularly central and factor-pure facet of Agreeableness, one that focuses on the domain's affective component (Goldberg, 1999; Hofstee et al., 1992). Politeness or Respectfulness (treating others with regard for their personal preferences and rights, while inhibiting antagonistic and aggressive impulses) highlights a key behavioral facet of Agreeableness (DeYoung et al., 2007), whereas interpersonal trust (holding positive generalized beliefs about others) represents an important cognitive facet of this domain (Goldberg, 1999; McCrae & Costa, 2008). We therefore selected Compassion, Respectfulness, and Trust as the BFI-2 Agreeableness facets.

Previous research has proposed several facet-level models of Conscientiousness (e.g., MacCann et al., 2009; Roberts et al., 2004; Roberts et al., 2005). Despite their differences, these models have consistently included three key facets. Organization or orderliness (preference for Order and structure) is a factor-pure, largely inhibitory facet of Conscientiousness (Hofstee et al., 1992; Saucier & Ostendorf, 1999). Productiveness or industriousness (work ethic and persistence while pursuing goals) captures a more proactive facet of Conscientiousness, and helps explain why this domain is a powerful predictor of academic achievement and job performance (Barrick & Mount, 1991; Duckworth, Peterson, Matthews, & Kelly, 2007; Poropat, 2009). Finally, Responsibility or reliability (commitment to meeting duties and obligations) represents the prosocial component of Conscientiousness, by capturing the degree to which a person can be depended upon by others (McCrae & Costa, 2008; Roberts et al., 2004, 2005). We therefore selected Organization, Productiveness, and Responsibility as the BFI-2 Conscientiousness facets.

Negative Emotionality represents individual differences in the frequency and intensity of negative affect (Clark & Watson, 2008; Widiger, 2009). Following research on basic emotions (Ekman, 1992), facet-level models of this domain typically distinguish

Table 1
The BFI-2 Facets Aligned With Previous Hierarchical Big Five Models

BFI-2 domains and facets	NEO PI-R facets (McCrae & Costa, 2008)	AB5C facets (Goldberg, 1999; Hofstee et al., 1992)	Lexical subcomponents (Saucier & Ostendorf, 1999)	Big Five aspects (DeYoung et al., 2007)
Extraversion				
Sociability	Gregariousness	Gregariousness	Sociability	Enthusiasm
Assertiveness	Assertiveness	Assertiveness	Assertiveness	Assertiveness
Energy Level	Positive Emotions/Activity	—	Activity-adventurousness	Enthusiasm
Agreeableness				
Compassion	Altruism	Understanding	Warmth-affection	Compassion
Respectfulness	Compliance	Cooperation	Gentleness	Politeness
Trust	Trust	Pleasantness	—	—
Conscientiousness				
Organization	Order	Orderliness	Orderliness	Orderliness
Productiveness	Self-Discipline	Efficiency	Industriousness	Industriousness
Responsibility	Dutifulness	Dutifulness	Reliability	—
Negative Emotionality				
Anxiety	Anxiety	Toughness (R)	Emotionality	Withdrawal
Depression	Depression	Happiness (R)	Insecurity	Withdrawal
Emotional Volatility	Angry Hostility	Stability (R)	Irritability	Volatility
Open-Mindedness				
Intellectual Curiosity	Ideas	Intellect	Intellect	Intellect
Aesthetic Sensitivity	Aesthetics	Reflection	—	Openness
Creative Imagination	Fantasy	Ingenuity	Imagination-creativity	—

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

between types of negative affect, especially fear/anxiety, sadness/depression, and irritation/anger (Goldberg, 1999; McCrae & Costa, 2008; Saucier & Ostendorf, 1999). Of these, the tendency to experience anxiety and fear tends to be most central and factor-pure (Hofstee et al., 1992). The tendency toward Depression and sadness is often accompanied by low levels of energy and arousal, and thus low Extraversion, whereas volatile mood swings often disrupt social interactions, and thus relate with low Agreeableness (Goldberg, 1999; McCrae & Costa, 2008). We therefore selected Anxiety, Depression, and Emotional Volatility as the BFI-2 Negative Emotionality facets.

Despite overall consensus around the Big Five model, there has been considerable disagreement regarding the defining features and optimal label for its fifth and final domain (DeYoung, 2014). Some researchers prefer the label Intellect, defined by intellectual interests and enjoyment of thinking (e.g., Goldberg, 1999), some prefer Imagination, defined by creativity and originality (e.g., Saucier, 1992), and others prefer Openness to Experience as a broader alternative that is primarily defined by intellectual and artistic interests but also includes a number of other characteristics (e.g., McCrae, 1994). We therefore selected Intellectual Curiosity, Creative Imagination, and Aesthetic Sensitivity as the three BFI-2 Open-Mindedness facets without a clear expectation that one of these would be uniquely factor-pure.

Phase 2: Identifying potential item content. After selecting and defining the 15 facet traits described above, our next step was to identify potential item content relevant to each facet. We did this using data from the ESCS. Specifically, from the list of 885 trait-descriptive adjectives administered to this sample, we identified one to three marker adjectives that seemed particularly central to the definition of each facet and did not correlate too strongly with potential markers of the other same-domain facets. For example, within the Extraversion domain, we selected "outgoing," "sociable," and "social" as markers of Sociability, "assertive" as a marker of Assertiveness, and "active," "energetic," and "enthusiastic" as markers of Energy Level. We then examined the correlations of these marker adjectives and the original BFI domain scales with the remaining trait-descriptive adjectives, the 2,552 IPIP items, and the 44 original BFI items. For each facet, we noted adjectives and items that correlated strongly and uniquely with the marker adjectives and BFI domain scales as potential sources of content for the BFI-2 item pool. For example, the IPIP item "Act as a leader" correlated .39 with the BFI Extraversion scale and .36 with the Assertiveness marker "assertive," compared with its next-strongest domain-level and facet-level correlations of .25 with BFI Conscientiousness and .26 with "outgoing." These correlations indicate that "Act as a leader" shares more than twice as much variance with Extraversion as with any other Big Five domain (i.e., $.39^2/.25^2 = 2.43$), and about twice as much variance with Assertiveness as with either of the other proposed Extraversion facets (i.e., $.36^2/.26^2 = 1.92$). Similarly, the trait-descriptive adjective "dominant" correlated .35 with BFI Extraversion and .46 with "assertive," compared with its next-strongest correlations of −.25 with BFI Agreeableness and .17 with "energetic." "Act as a leader" and "dominant" were therefore noted as potential sources of Assertiveness content.

Phase 3: Constructing a preliminary item pool. Drawing on the results of Phase 2, we constructed a preliminary pool of 110 potential BFI-2 items. This pool consisted of the 44 original BFI items, revised versions of 19 BFI items, and 47 new items. The 19 revised versions included original BFI items edited to (a) clarify their associations with the Big Five domains (e.g., the item "Can be cold and aloof" was revised to "Can be cold and uncaring" to strengthen its association with Agreeableness and weaken its association with Extraversion), (b) clarify their associations with the BFI-2 facets (e.g., the item "Is a reliable worker" was revised to "Is reliable, can always be counted on" to strengthen its association with Responsibility and weaken its association with Productiveness), and (c) make them easier to understand by elaborating their meaning (e.g., the item "Is inventive" was revised to "Is inventive, finds clever ways to do things") or replacing difficult words (e.g., the item "Values artistic, aesthetic experiences" was revised to "Values art and beauty"). The 47 new items were constructed using content from the trait-descriptive adjectives and IPIP items identified in Phase 2. For example, the adjective "dominant" and the IPIP item "Act as a leader" were combined to form the candidate Assertiveness item "Is dominant, acts as a leader." We ensured that the preliminary pool included at least three true-keyed and three false-keyed items for each BFI-2 facet, to facilitate the construction of content-balanced scales that would automatically control for individual differences in acquiescence.

Conclusion

Study 1 yielded two key outcomes. The first was a hierarchical model of personality structure, with 15 facet traits nested within the Big Five domains (see Table 1). This model identifies conceptually and empirically prominent Big Five facets from previous research, and organizes them within an integrative hierarchical structure. We therefore propose that these traits constitute a minimally necessary set for capturing facet-level personality information within each Big Five domain. The second key outcome of Study 1 was a broad, deep, and balanced pool of 110 potential BFI-2 items whose content reflects this intended hierarchical structure.

Study 2

Study 2 was conducted to refine the preliminary item pool from Study 1 into the final BFI-2. To do this, we administered the preliminary item pool to a large and diverse sample of adults. We then used these data to select the final set of BFI-2 items, organize these items into facet and domain scales, and conduct a preliminary examination of the BFI-2's basic measurement properties and multidimensional structure.

Method

Participants and procedure. Participants in this study were 1,000 adult residents of English-speaking nations who rated themselves using the preliminary BFI-2 item pool. They did this at personalitylab.org, a website that offers its visitors free, anonymous feedback on a number of personality-related questionnaires. To balance gender effects while constructing the final BFI-2 scales, our final sample of 1,000 participants was randomly selected from a larger pool of respondents to include 500 men and 500 women. Members of the final sample ranged in age from 18 to 77 years old ($M = 29.25$, $SD = 12.17$), with most (64%) under

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the age of 30. They were diverse in terms of ethnicity: 65% described themselves as White/Caucasian, 7% as Hispanic/Latino, 7% as Black/African American, 6% as Asian/Asian American, 1% as Native American/American Indian, 4% as another ethnicity, and 5% as mixed ethnicity, with 5% not reporting ethnicity. Most participants (79%) were residents of the United States, with smaller numbers residing in the United Kingdom (9%), Canada (7%), and Australia or New Zealand (5%).

Measure. All participants described themselves using the complete pool of 110 candidate BFI-2 items developed in Study 1. They rated these items using a 5-point scale ranging from *disagree strongly* to *agree strongly*. As a preliminary data check, we computed alpha reliabilities for the original BFI domain scales in the present sample. Alphas were .87 for Extraversion, .82 for Agreeableness, .84 for Conscientiousness, .86 for Negative Emotionality, and .82 for Open-Mindedness. These are essentially identical to the corresponding values from the ESCS data analyzed in Study 1, and very similar to previously published results from other noninternet samples (John et al., 2008).

Results and Discussion

As discussed above, our major goals in developing the BFI-2 were to retain the focus, clarity, and brevity of the original BFI, to develop a robust hierarchical structure, to balance bandwidth and fidelity, and to minimize the influence of acquiescence. Our rational and empirical criteria for constructing the final BFI-2 scales reflected these goals. To ensure brevity, we limited the BFI-2 to four items per facet. To promote focus and clarity, we rationally evaluated item content and readability. To promote a robust hierarchical structure, we examined the BFI-2's multidimensional structure at the domain and facet levels. To balance bandwidth and fidelity, we considered the items' convergent and discriminant correlations with preliminary domain and facet scales, as well as rational judgments of item redundancy. Finally, to minimize the influence of acquiescence, we included an equal number of true-keyed and false-keyed items on each scale. We used these criteria to select a final set of 60 items for the BFI-2, with 15 4-item facet scales that aggregate into five 12-item domain scales. The BFI-2 self-report form, as well as scoring keys for the domain and facet scales, are presented in the Appendix.

We next examined the BFI-2's basic measurement properties and domain-level structure in the scale development sample. Alpha reliabilities for the 12-item domain scales were .87 for Extraversion, .83 for Agreeableness, .88 for Conscientiousness, .91 for Negative Emotionality, and .84 for Open-Mindedness. Alphas for the 4-item facet scales ranged from .65 to .84, with a mean of .76. In a principal components analysis (PCA) of the 15 BFI-2 facets, when we extracted and varimax rotated five components, all 15 facets had their strongest loading on the intended component, and all of these primary loadings were at least .73 in strength (M = .80). In a similar PCA of the BFI-2 items, all 60 items had their strongest loading on the intended component, and all of these primary loadings were at least .41 in strength (M = .60).

These preliminary results suggest that our joint rational-empirical approach to developing the BFI-2 yielded a reliable measure with a clear Big Five structure. However, these results may be favorably biased by inadvertent capitalization on chance, in that they are computed from the same scale development sample used to refine the preliminary BFI-2 item pool into the final measure. We therefore refrain from reporting additional analyses of this sample. Instead, we evaluate the BFI-2 more thoroughly using data from independent validation samples.

Study 3

Study 3 was conducted to examine the BFI-2's basic measurement properties, multidimensional structure, and nomological network, using two independent samples: an Internet sample and a student sample. Consistent with our goals for developing the BFI-2, we were particularly interested in examining (a) basic measurement properties, including internal consistency, retest reliability, and self-peer agreement, (b) multidimensional structure across levels of analysis (facets within domains, items within domains, and items within facets), (c) the influence of acquiescence on item responses, (d) convergence with other Big Five measures, and (e) substantive relations with and capacity to predict self-reported and peer-reported criteria.

Method

Participants and procedure.

Internet validation sample. These participants were 1,000 adult visitors to personalitylab.org (500 men and 500 women) who rated themselves using the original BFI and the BFI-2. The procedures used to recruit, select, and assess these participants were identical to those used in Study 2, but the scale development and Internet validation samples did not overlap with each other. Participants in the Internet validation sample ranged in age from 18 to 74 years old (M = 28.73, SD = 11.68), with most (65%) under the age of 30. Regarding ethnicity, 66% described themselves 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 mixed ethnicity, 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%).

Student validation sample. These participants were 470 students (146 men, 313 women, and 11 who did not report gender) enrolled in psychology courses at a large, public university who completed the BFI and BFI-2. They ranged in age from 16 to 49 years old (M = 21.68, SD = 3.26), with most (90%) under the age of 25. Regarding ethnicity, 49% described themselves as Asian/Asian American, 27% as White/Caucasian, 12% as Hispanic/Latino, 1% as Black/African American, and 8% as another ethnicity, with 4% not reporting ethnicity. Each participant completed a series of questionnaires at a course website in exchange for partial course credit.

Measures. All participants described themselves using the combined BFI and BFI-2 item set, allowing us to score both measures. In the Internet validation sample, alpha reliabilities for the original BFI domain scales were .88 for Extraversion, .79 for Agreeableness, .85 for Conscientiousness, .85 for Negative Emotionality, and .81 for Open-Mindedness, and alphas for the post hoc BFI facet scales ranged from .54 to .82 with a mean of .69. In the student validation sample, alphas for the BFI domain scales were .88 for Extraversion, .80 for Agreeableness, .81 for Consci-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

entiousness, .84 for Negative Emotionality, and .80 for Open-Mindedness, and alphas for the post hoc facet scales ranged from .54 to .83 with a mean of .69. These results are similar to previously published values (e.g., John et al., 2008; Soto & John, 2009a). To examine the BFI-2's reliability and nomological network, participants in the student validation sample were also assessed using a number of additional measures, described below.

Big Five Mini-Markers. Within one week of completing the BFI and BFI-2, 438 student participants completed four additional measures of the Big Five: the Big Five Mini-Markers, the Big Five Aspect Scales, the NEO Personality Inventory—Revised, and the NEO Five-Factor Inventory. Completed during the same session as the BFI and BFI-2, the Mini-Markers (Saucier, 1994) are a 40-item short form of Goldberg's (1992) unipolar Big Five marker adjectives. The items are trait-descriptive adjectives that respondents rate on a 9-point scale ranging from *extremely inaccurate* to *extremely accurate* as a description of the respondent. In this sample, alpha reliabilities for the Mini-Marker scales were .87 for Extraversion, .84 for Agreeableness, .84 for Conscientiousness, .84 for Negative Emotionality, and .82 for Open-Mindedness.

Big Five Aspect Scales. Also completed during the same session as the BFI and BFI-2, the Big Five Aspect Scales (BFAS; DeYoung et al., 2007) are a hierarchical, 100-item measure developed from the IPIP. The BFAS items are descriptive phrases that respondents rate on a 5-point scale ranging from *strongly disagree* to *strongly agree*. The BFAS uses 10-item scales to assess two aspect traits within each Big Five domain. Each Big Five domain is then scored by aggregating its two aspects. In this sample, alpha reliabilities for the BFAS domains were .89 for Extraversion, .87 for Agreeableness, .85 for Conscientiousness, .91 for Negative Emotionality, and .84 for Open-Mindedness. Alphas for the 10 aspect scales ranged from .76 to .90, with a mean of .84.

NEO Personality Inventory—Revised and NEO Five-Factor Inventory. Completed approximately one week after the BFI and BFI-2, the NEO Personality Inventory—Revised (NEO PI-R; McCrae & Costa, 2008) is a 240-item, hierarchical Big Five measure. Its items are statements that respondents rate on a 5-point scale ranging from *strongly disagree* to *strongly agree*. The NEO PI-R uses 8-item scales to assess six facet traits within each Big Five domain, then uses factor scoring equations to score the five superordinate domains from the 30 facets. In this sample, alpha reliabilities for the NEO PI-R domains were .90 for Extraversion, .89 for Agreeableness, .92 for Conscientiousness, .93 for Negative Emotionality, and .89 for Open-Mindedness. Alphas for the 30 facet scales ranged from .57 to .84, with a mean of .74.

We also scored the five domain scales of the NEO Five-Factor Inventory (NEO-FFI; McCrae & Costa, 2008) from a subset of 60 NEO PI-R items. The NEO-FFI is a short form of the NEO PI-R that assesses each Big Five domain using a 12-item scale. In this sample, alphas for the NEO-FFI domain scales were .81 for Extraversion, .78 for Agreeableness, .85 for Conscientiousness, .87 for Negative Emotionality, and .73 for Open-Mindedness.

Behavioral self-reports. Approximately two weeks after completing the BFI and BFI-2, 439 student participants 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*. After within-person centering each participant's set of ratings to control for acquiescence (Bardi & Schwartz, 2003), the items were aggregated into 10 scales corresponding with values from the Schwartz values circumplex: conformity (e.g., "Obey rules and regulations"), tradition (e.g., "Attend religious/spiritual services on important dates"), benevolence (e.g., "Lend things to people I know"), power (e.g., "Manipulate others to get what I want"), universalism (e.g., "Sign petitions to support environmental protection efforts"), hedonism (e.g., "Indulge myself by buying things that I don't really need"), security (e.g., "Go out of my way not to catch colds, the flu, etc. from others"), stimulation (e.g., "Look for stimulating activities that break up my routine"), achievement (e.g., "Study hard in order to get the highest grade in class"), and self-direction (e.g., "Read about intellectual topics that are not directly related to my classes or professional goals"). As in previous research, and reflecting each value's broad range of conceptually relevant behaviors, alpha reliabilities varied considerably across the scales, from .27 to .70 (cf. Bardi & Schwartz, 2003; Pozzebon & Ashton, 2009). While certainly noting the low internal consistency of some scales, we retained their original scoring so that they could serve as an unbiased, a priori set of criteria for examining the BFI-2's nomological network.

Psychological Well-Being Scales. Approximately two weeks before completing the BFI and BFI-2, 179 student participants completed the Psychological Well-Being Scales (Ryff, 1989). This measure includes 84 items that respondents rate on a 5-point scale ranging from *disagree strongly* to *agree strongly*. The items are aggregated into 14-item scales assessing six aspects of psychological well-being: positive relations with others (e.g., "I feel like I get a lot out of my friendships"), purpose in life (e.g., "I have a sense of direction and purpose in life"), environmental mastery (e.g., "I am quite good at managing the many responsibilities of my daily life"), self-acceptance (e.g., "For the most part, I am proud of who I am and the life I lead"), autonomy (e.g., "My decisions are not usually influenced by what everyone else is doing"), and personal growth (e.g., "I have the sense that I have developed a lot as a person over time"). In this sample, the scales' alpha reliabilities ranged from .85 to .92.

Retest assessment and peer-reports. Approximately eight weeks after completing the BFI and BFI-2, a subsample of 110 student participants completed the BFI-2 a second time, and a separate subsample of 184 student participants were described by a knowledgeable peer. Most peers were friends (60%) or romantic partners (30%). Each peer rated the target participant using the BFI-2, as well as items assessing four criteria: social connectedness (4 items, e.g., "Has close relationships with others"), likability (2 items, e.g., "Is the kind of person almost everyone likes"), stress resistance (5 items, e.g., "Is able to put emotionally difficult events into proper perspective"), and positive affect (2 items, e.g., "Is joyful, happy, pleased"; English, John, Srivastava, & Gross, 2012). Alpha reliabilities for the peer-reported criteria were .76 for social connectedness, .75 for likability, .78 for stress resistance, and .67 for positive affect.

Results and Discussion

Reliability, self-peer agreement, and descriptive statistics. Our first objective for Study 3 was to examine the BFI-2's basic measurement properties in the Internet and student validation samples. We expected the BFI-2 to demonstrate strong internal consistency, retest reliability, between-domain discrimination, and

self-peer agreement. Tables 2 and 3 present reliability coefficients and interscale correlations for the BFI-2 domain scales and facet scales, respectively. As shown in Table 2, the five domain scales all had alpha reliabilities of at least .83 in the Internet validation sample and .85 in the student sample, as well as retest reliabilities of at least .76 in the student sample. As shown in Table 3, alpha reliabilities for the 15 facet scales ranged from .66 to .85 in each sample, averaging .76 in the Internet sample and .77 in the student sample; these scales' retest reliabilities in the student sample ranged from .66 to .83, averaging .73. These results indicate high reliability for the BFI-2 domain scales, as well as adequate to high reliability for the 4-item facet scales.

Tables 2 and 3 also present interscale correlations among the BFI-2 domain and facet scales, respectively. Absolute correlations between the domain scales averaged only .20 in the Internet validation sample and .24 in the student validation sample. As for the facet scales, within-domain correlations (e.g., between Sociability and Assertiveness) averaged .55 in the student validation sample and .53 in the Internet validation sample, whereas absolute between-domain correlations (e.g., between Sociability and Intellectual Curiosity) averaged only .16 in the Internet validation sample and .19 in the student validation sample. These results indicate strong discrimination between the BFI-2 domains, as well as an appropriately moderate degree of discrimination among the facets within each domain.

Self-peer agreement correlations for the BFI-2 domain and facet scales in the student validation sample are presented in Table 4. At the domain level, self-peer agreement correlations ranged from .42 to .69, and averaged .56; in contrast, absolute discriminant correlations averaged only .11. At the facet level, agreement correlations ranged from .27 to .73, and averaged .49; in contrast, within-domain discriminant correlations averaged only .32, and absolute between-domain discriminant correlations averaged only .10. Moreover, 13 of the 15 self-reported facets had their strongest correlation with the corresponding peer-reported facet; conversely, 14 of the 15 peer-reported facets had their strongest correlation with the corresponding self-reported facet. These findings indicate that the BFI-2 shows strong self-peer agreement and discrimination at both the domain and facet levels.

Table 5 presents descriptive statistics for the BFI-2 domain and facet scales, as well as the mean-level gender differences within each sample and mean-level differences between the two samples. Converging with previous research using other measures (e.g., Costa, Terracciano, & McCrae, 2001), women tended to describe themselves as somewhat more extraverted, agreeable, conscientious, and emotional than men. However, some facet traits showed distinctive gender differences. For example, within the Extraversion domain, women tended to report greater Sociability and Energy Level—but not greater Assertiveness—than men. Within the Open-Mindedness domain, women tended to report greater Aesthetic Sensitivity, whereas men tended to report greater Creative Imagination and Intellectual Curiosity. As for comparisons between the two samples, members of the Internet sample reported somewhat higher levels of Open-Mindedness and Negative Emotionality (especially Depression and Emotional Volatility), as well as somewhat lower levels of Organization, than did members of the student sample. This pattern seems consistent with intuitions about individuals who might be especially likely to seek out personality feedback online.

Multidimensional structure and influence of acquiescence.

Domain-level structure. Our second and third objectives for Study 3 were to examine (a) the BFI-2's multidimensional structure at the domain and facet levels, and (b) the influence of individual differences in acquiescence on this structure. At the domain level, we expected that there would be some influence of acquiescent responding, but that the BFI-2 would still show a robust Big Five structure. As an initial test for acquiescence effects, we compared results from PCAs of the 60 raw BFI-2 items with results from PCAs of the same 60 items after within-person centering to control for acquiescence (i.e., after computing each participant's mean response across the full set of 60 items and then subtracting this within-person mean from each of their individual item responses). Figure 1 presents the eigenvalues of the unrotated components from each analysis. Reflecting the expected influence of acquiescence, Figures 1a and 1c show that, in both samples, PCAs of the raw items indicated six components above the scree line: the Big Five plus an additional acquiescence component (cf. Rammstedt & Farmer, 2013; Soto et al., 2008). However, as shown in Figures 1b and 1d, within-person centering to control for acquiescence effectively eliminated the additional component.

To test whether the BFI-2's intended domain-level structure would emerge despite the influence of acquiescence, we next extracted and varimax rotated five principal components from the raw items and centered items in each sample; loadings from analyses of the centered items are presented in Table 6. Each analysis produced a clear Big Five structure, with each item

Table 2
Reliability Coefficients and Intercorrelations of the BFI-2 Domains (Study 3)

	Reliability		Intercorrelations			
Domain	Alpha	Retest	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality
Extraversion	.88/.88	.84				
Agreeableness	.83/.85	.76	.14/.09			
Conscientiousness	.88/.86	.83	.22/.27	.28/.28		
Negative Emotionality	.90/.90	.81	−.34/−.36	−.29/−.31	−.30/−.34	
Open-Mindedness	.84/.85	.76	.20/.21	.15/.26	−.02/.14	−.06/−.10
Mean	***.87/.87***	***.80***				

Note. Values left of each forward slash are for the Internet validation sample (N = 1,000); values right of each forward slash are for the student validation sample (N = 470). Retest = Eight-week retest reliability in the student validation sample (N = 110). Absolute correlations of .07 or stronger in the Internet sample, or .10 or stronger in the student sample, are significant at $p < .05$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Reliability Coefficients and Intercorrelations of the BFI-2 Facets (Study 3)

Facet	Reliability		Extraversion			Agreeableness			Conscientiousness			Negative Emotionality			Open-Mindedness	
	Alpha	Retest	Soc.	Ass.	Ene.	Com.	Res.	Tru.	Org.	Pro.	Resy.	Anx.	Dep.	Emo.	Int.	Aes.
Sociability	.85/.85	.83	—													
Assertiveness	.78/.78	.80	**.54/.59**	—												
Energy Level	.76/.72	.74	**.59/.63**	**.48/.44**	—											
Compassion	.66/.72	.68	.12/.08	.03/.03	.29/.28	—										
Respectfulness	.71/.72	.66	−.05/−.16	−.14/−.19	.15/.07	**.58/.51**	—									
Trust	.70/.74	.75	.16/.15	−.01/.03	.28/.25	**.48/.53**	**.50/.50**	—								
Organization	.85/.84	.76	−.01/−.01	.16/.15	.13/.15	.09/.17	.20/.23	.06/.06	—							
Productiveness	.78/.73	.74	.14/.22	.31/.36	.35/.37	.22/.18	.26/.16	.15/.14	**.60/.52**	—						
Responsibility	.74/.66	.68	.04/.06	.17/.26	.20/.25	.30/.32	.37/.32	.20/.22	**.52/.50**	**.60/.58**	—					
Anxiety	.78/.79	.79	−.20/−.22	−.21/−.21	−.23/−.21	−.02/.00	−.16/−.09	−.31/−.30	−.05/−.05	−.16/−.14	−.16/−.12	—				
Depression	.84/.83	.74	−.39/−.36	−.39/−.35	−.47/−.52	−.14/−.20	−.18/−.18	−.34/−.36	−.21/−.23	−.36/−.42	−.31/−.39	**.64/.56**	—			
Emotional Volatility	.84/.84	.70	−.06/−.11	−.12/−.17	−.15/−.19	−.08/−.12	−.30/−.34	−.32/−.31	−.17/−.20	−.27/−.33	−.31/−.30	**.65/.58**	**.59/.60**	—		
Intellectual Curiosity	.67/.73	.78	.00/.10	.18/.29	.10/.20	.12/.27	.04/.16	.04/.11	−.10/.10	.01/.18	−.04/.21	.03/−.07	−.01/−.12	−.02/−.09	—	
Aesthetic Sensitivity	.76/.83	.67	−.02/−.03	.11/.06	.11/.05	.14/.28	.09/.18	.09/.09	−.07/.02	−.01/−.01	−.05/.06	.02/.04	.03/.02	.02/.05	**.49/.44**	—
Creative Imagination	.75/.78	.67	.18/.22	.33/.26	.29/.24	.16/.17	.08/.17	.12/.12	−.02/.02	.14/.14	.05/.18	−.13/−.23	−.21/−.16	−.12/−.13	**.48/.48**	**.44/.44**
Mean	***.76/.77***	***.73***														

Note. Values left of each forward slash are for the Internet validation sample (N = 1,000); values right of each forward slash are for the student validation sample (N = 470). Retest = Eight-week retest reliability in the student validation sample (N = 110). Soc. = Sociability; Ass. = Assertiveness; Ene. = Energy Level; Com. = Compassion; Res. = Respectfulness; Tru. = Trust; Org. = Organization; Pro. = Productiveness; Resy. = Responsibility; Anx. = Anxiety; Dep. = Depression; Emo. = Emotional Volatility; Int. = Intellectual Curiosity; Aes. = Aesthetic Sensitivity. Within-domain correlations are bolded. Absolute correlations of .07 or stronger in the Internet sample, or .10 or stronger in the student sample, are significant at $p < .05$.

loading most strongly on its intended domain component in both samples. For the raw items, these primary loadings were all .37 or stronger (M = .61) in the Internet validation sample, and .41 or stronger (M = .60) in the student validation sample. For the centered items, the primary loadings were all .39 or stronger (M = .62) in the Internet sample, and .45 or stronger (M = .61) in the student sample. By comparison, for both the raw and centered items the absolute secondary loadings averaged only .10 in the Internet sample and .12 in the student sample. Moreover, Congruence coefficients comparing pairs of corresponding components between the two samples were all .97 or greater, indicating near-perfect replication.

To further examine the BFI-2's domain-level structure, we conducted a PCA of the 15 facet scales (scored from the raw item responses) in each sample. In both samples, the pattern of eigenvalues indicated only five components above the scree line. For example, in the Internet validation sample the first five eigenvalues were 4.05, 2.16, 1.90, 1.60, and 1.32, followed by 0.56, 0.49, 0.48, 0.44, and 0.41. This result indicates that the content-balanced facet scales effectively control for acquiescence, and therefore do not produce an additional acquiescence component beyond the Big Five.

We next extracted and varimax rotated five components from the facet scales in each sample. The loadings from these analyses, presented in Table 7, reveal three noteworthy findings. First, each facet loaded strongly on its intended Big Five component; these 15 primary loadings were all at least .73 (M = .81) in the Internet validation sample, and at least .67 (M = .79) in the student validation sample. Second, several facets had conceptually meaningful patterns of secondary loadings. For example, Depression loaded negatively on Extraversion, capturing the association of sadness with lack of energy and social confidence. Similarly, the negative loading of Trust on Negative Emotionality reflects the association of holding negative beliefs about others with experiencing unpleasant emotions in social situations. Third, the overall pattern of facet loadings replicated very clearly between the Internet and student samples; Congruence coefficients comparing pairs of corresponding components were all .98 or greater.

Facet-level structure. We expected the BFI-2 to show a robust multidimensional structure not only at the domain level, but also at the facet level within each domain. To test this hypothesis, and to further investigate the influence of acquiescence, we fit a series of five confirmatory factor analysis (CFA) models to the raw items within each Big Five domain in each sample. Fit statistics for these models are presented in Table 8. The initial, *single domain* model simply allowed all 12 items within a domain to load on a single factor. As expected, this model fit poorly for each domain in both samples (CFIs ≤ .805 in the Internet validation sample and .800 in the student validation sample).[1] This finding suggests the presence of multiple dimensions within each domain, and provides a baseline for evaluating more-complex models.

Each of the next two models allowed for individual differences in acquiescence. Specifically, the bifactor *single domain plus acquiescence* model allowed each item within a Big Five domain to load on both the substantive domain factor and an acquiescence method

[1] In our description of results, we focus on the comparative fit index (CFI) to summarize model fit. However, all of the fit statistics presented in Table 8 support similar conclusions.

Table 4
Self–Peer Agreement Correlations for the BFI-2 Domains and Facets (Study 3)

	Peer-reported domains					Peer-reported facets														
Report	Ext.	Agr.	Con.	Neg.	Ope.	Soc.	Ass.	Ene.	Com.	Res.	Tru.	Org.	Pro.	Resy.	Anx.	Dep.	Emo.	Int.	Aes.	Cre.
Self-reported domains																				
Extraversion	**.69**	.01	.04	−.28	−.05	.69	.55	.44	.09	−.18	.11	−.05	.16	.00	−.27	−.37	−.10	.02	−.17	.04
Agreeableness	.05	**.51**	.17	−.13	.18	.09	−.16	.18	.43	.40	.44	.05	.18	.19	−.06	−.13	−.14	.17	.13	.12
Conscientiousness	.14	.13	**.53**	−.15	.12	.04	.16	.17	.13	.07	.13	.46	.45	.33	−.03	−.21	−.15	.12	.10	.06
Negative Emotionality	−.21	−.17	−.11	**.63**	.05	−.19	−.10	−.21	.00	−.07	−.31	−.05	−.18	−.03	.55	.54	.55	.01	.14	−.05
Open-Mindedness	.04	.07	.10	.03	**.42**	.05	.01	.02	.11	.05	.03	.01	.16	.09	.05	.03	.00	.33	.35	.32
Self-reported facets																				
Sociability	.65	.02	−.03	−.23	−.10	**.73**	.46	.40	.07	−.15	.11	−.09	.08	−.06	−.26	−.30	−.06	.03	−.23	−.02
Assertiveness	.51	−.08	.03	−.19	−.08	.47	**.56**	.23	−.01	−.22	.03	−.06	.14	.01	−.19	−.24	−.07	−.04	−.13	−.02
Energy Level	.56	.11	.13	−.29	.06	.52	.36	**.50**	.18	−.06	.14	.04	.20	.06	−.24	−.41	−.13	.06	−.06	.17
Compassion	.04	.37	.14	.02	.21	.07	−.10	.13	**.41**	.26	.26	.01	.16	.18	.04	.01	.00	.16	.17	.18
Respectfulness	−.10	.42	.15	−.13	.10	−.10	−.21	.07	.30	**.41**	.32	.05	.15	.16	−.02	−.10	−.19	.15	.05	.03
Trust	.16	.49	.13	−.21	.13	.23	−.09	.24	.36	.34	**.50**	.06	.13	.13	−.17	−.22	−.16	.11	.10	.09
Organization	.05	.12	.48	−.14	.06	−.02	.07	.08	.11	.09	.11	**.56**	.26	.27	−.04	−.19	−.13	.05	.07	.02
Productiveness	.22	.06	.41	−.16	.09	.11	.22	.24	.06	.02	.07	.25	**.51**	.22	−.07	−.19	−.15	.14	.04	.04
Responsibility	.07	.14	.39	−.06	.17	.00	.10	.10	.16	.06	.13	.26	.35	**.32**	.06	−.13	−.08	.13	.15	.11
Anxiety	−.16	−.08	.04	.56	.08	−.19	−.04	−.16	.09	.01	−.26	.05	−.04	.09	**.57**	.42	.47	.06	.15	−.04
Depression	−.29	−.09	−.13	.46	.07	−.29	−.19	−.23	.00	.03	−.22	−.06	−.18	−.08	.39	**.49**	.33	.03	.15	−.02
Emotional Volatility	−.08	−.24	−.17	.56	−.01	−.02	−.03	−.15	−.09	−.19	−.30	−.10	−.22	−.08	.44	.45	**.57**	−.05	.07	−.06
Intellectual Curiosity	.07	.10	.14	.08	.46	.06	.05	.07	.13	.07	.05	.02	.19	.13	.12	.07	.03	**.49**	.26	.35
Aesthetic Sensitivity	−.09	.09	.13	.03	.32	−.06	−.06	−.10	.10	.10	.02	.09	.12	.11	.05	.06	−.03	.16	**.42**	.17
Creative Imagination	.13	−.01	−.02	−.03	.23	.15	.05	.11	.02	−.06	.00	−.10	.07	−.01	−.05	−.05	.01	.17	.10	**.27**

Note. $N = 184$. Absolute correlations of .15 or stronger are significant at $p < .05$. Ext. = Extraversion; Agr. = Agreeableness; Con. = Conscientiousness; Neg. = Negative Emotionality; Ope. = Open-Mindedness; Soc. = Sociability; Ass. = Assertiveness; Ene. = Energy Level; Com = Compassion; Res. = Respectfulness; Tru. = Trust; Org. = Organization; Pro. = Productiveness; Resy. = Responsibility; Anx. = Anxiety; Dep. = Depression; Emo. = Emotional Volatility; Int. = Intellectual Curiosity; Aes. = Aesthetic Sensitivity; Cre. = Creative Imagination. Self-peer agreement correlations are bolded.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

factor. All loadings on the acquiescence factor were constrained to equal 1 (thereby forcing true-keyed and false-keyed items to load in the same direction), and the acquiescence factor was not allowed to correlate with the domain factor; these constraints ensured that the acquiescence factor would represent individual differences in response style that were distinct from meaningful personality content (cf. Billiet & McClendon, 2000; John et al., 2008). The *positive and negative items* model included two correlated factors: one defined by the domain's true-keyed items, and one defined by its false-keyed items. As expected, fit statistics for the *single domain plus acquiescence* and *positive and negative items* models were essentially equivalent to each other. Compared with the baseline *single domain* model, they provided a substantial improvement in fit for Agreeableness (ΔCFIs $\geq$ .070 in the Internet sample and .047 in the student sample), and more modest improvements for the other four domains (ΔCFIs $\leq$ .024 in the Internet sample and .021 student sample). These results indicate that divergence between the high and low pole of each domain can be explained by individual differences in acquiescence, and that acquiescence has a particularly strong influence on responses to Agreeableness items. However, neither model provided acceptable fit for any domain (CFIs $\leq$ .876 in the Internet sample and .850 in the student sample), suggesting the presence of additional dimensions within each domain.

The fourth, *three facets* model included three factors representing the three BFI-2 facet scales within a Big Five domain. Each item was only allowed to load on a single facet factor, and the three facet factors were allowed to intercorrelate. Compared with the *single domain* model, the *three facets* model provided a substantial improvement in fit for each domain in both samples (ΔCFIs $\geq$ .050 in the Internet sample and .076 in the student sample). Moreover, this model provided acceptable or nearly acceptable overall fit for Extraversion, Conscientiousness, Negative Emotionality, and Open-Mindedness in both samples (CFIs $\geq$ .902 in the Internet sample and .895 in the student sample).

The final, *three facets plus acquiescence* model (illustrated in Figure 2) added an acquiescence factor to the *three facets* model just described. Thus, each item was allowed to load on both its facet factor and an acquiescence method factor. All loadings on the acquiescence factor were constrained to equal 1, and this factor was not allowed to correlate with any of the facet factors. Compared with the *single domain* model, this final model provided a large improvement in fit for each Big Five domain in both samples (ΔCFIs $\geq$ .145 in the Internet sample and .149 in the student sample). Moreover, it provided acceptable overall fit for each domain in both samples (CFIs $\geq$ .930 in the Internet sample and .932 in the student sample).

Conclusions. Taken together, these results support three key conclusions. First, the BFI-2 has a robust hierarchical structure, with three distinguishable facets nested within each Big Five domain.[2] Second, acquiescence somewhat distorts this structure at both the domain level (by producing an additional acquiescence

[2] This conclusion was also supported by the results of exploratory PCAs. Specifically, when we extracted and promax-rotated three components from each Big Five domain's 12 within-person centered items, 58 of the 60 total items had their strongest loading on the expected facet component in each sample. Moreover, the two items with unexpected primary loadings were not consistent across the two samples.

Table 5
Descriptive Statistics for the BFI-2 Domains and Facets (Study 3)

Domain or facet	Internet sample Men *M* (*SD*)	Internet sample Women *M* (*SD*)	Internet sample Combined *M* (*SD*)	Internet sample Gender *d*	Student sample Men *M* (*SD*)	Student sample Women *M* (*SD*)	Student sample Combined *M* (*SD*)	Student sample Gender *d*	Sample *d*
Extraversion	3.15 (.78)	3.31 (.80)	3.23 (.80)	.21	3.20 (.70)	3.31 (.73)	3.25 (.71)	.15	−.03
Sociability	2.80 (1.02)	3.10 (1.07)	2.95 (1.05)	.29	2.94 (.86)	3.06 (1.01)	3.00 (.94)	.12	−.05
Assertiveness	3.28 (.92)	3.28 (.93)	3.28 (.93)	.01	3.27 (.82)	3.28 (.85)	3.28 (.84)	.02	.01
Energy Level	3.37 (.88)	3.56 (.89)	3.47 (.89)	.22	3.40 (.80)	3.58 (.72)	3.49 (.77)	.24	−.03
Agreeableness	3.57 (.65)	3.79 (.60)	3.68 (.64)	.35	3.51 (.63)	3.82 (.56)	3.66 (.62)	.53	.03
Compassion	3.72 (.79)	3.97 (.76)	3.84 (.78)	.33	3.60 (.81)	3.98 (.69)	3.79 (.78)	.49	.07
Respectfulness	3.87 (.73)	4.08 (.68)	3.98 (.71)	.30	3.76 (.68)	4.05 (.64)	3.91 (.68)	.44	.10
Trust	3.13 (.83)	3.32 (.80)	3.23 (.82)	.24	3.15 (.77)	3.43 (.77)	3.29 (.78)	.36	−.08
Conscientiousness	3.35 (.74)	3.50 (.79)	3.43 (.77)	.20	3.34 (.60)	3.54 (.66)	3.44 (.64)	.31	−.03
Organization	3.33 (.99)	3.51 (1.03)	3.42 (1.01)	.19	3.46 (.88)	3.68 (.87)	3.57 (.88)	.26	−.16
Productiveness	3.31 (.87)	3.43 (.93)	3.37 (.90)	.13	3.24 (.75)	3.39 (.80)	3.32 (.78)	.19	.07
Responsibility	3.40 (.78)	3.57 (.83)	3.48 (.81)	.20	3.33 (.60)	3.55 (.71)	3.44 (.66)	.33	.05
Negative Emotionality	2.95 (.88)	3.18 (.84)	3.07 (.87)	.27	2.84 (.74)	2.95 (.79)	2.89 (.76)	.14	.21
Anxiety	3.28 (.95)	3.58 (.88)	3.43 (.93)	.33	3.20 (.78)	3.53 (.85)	3.37 (.83)	.40	.07
Depression	2.82 (1.03)	2.88 (1.02)	2.85 (1.02)	.06	2.65 (.92)	2.53 (.93)	2.59 (.93)	−.14	.26
Emotional Volatility	2.77 (1.04)	3.09 (1.04)	2.93 (1.05)	.31	2.66 (.91)	2.79 (.97)	2.73 (.95)	.13	.20
Open-Mindedness	3.93 (.64)	3.91 (.67)	3.92 (.65)	−.02	3.71 (.65)	3.62 (.63)	3.66 (.64)	−.15	.39
Intellectual Curiosity	4.18 (.69)	4.03 (.71)	4.10 (.70)	−.21	3.89 (.76)	3.80 (.70)	3.85 (.73)	−.12	.24
Aesthetic Sensitivity	3.71 (.90)	3.88 (.94)	3.80 (.92)	.19	3.57 (.95)	3.58 (.90)	3.58 (.92)	.02	.36
Creative Imagination	3.89 (.81)	3.82 (.80)	3.85 (.81)	−.09	3.68 (.75)	3.46 (.77)	3.57 (.77)	−.28	.36
Sample size	500	500	1,000		313	146	459		

Note. Gender d = Cohen's d for the mean-level difference between men and women, with positive values indicating higher scores for women; differences of .13 or larger in the Internet sample, or .20 or larger in the student sample, are significant at $p < .05$. Sample d = Cohen's d for the mean-level difference between the combined-gender Internet and student samples, with positive values indicating higher scores in the Internet sample; differences of .12 or larger are significant at $p < .05$. Combined values in the student sample are estimated for a sample with an equal number of men and women.

component and suppressing the items' primary loadings) and the facet level (by biasing the associations between same-domain items; cf. John et al., 2008; McCrae et al., 2001; Rammstedt & Farmer, 2013; Soto et al., 2008). Third, the influence of acquiescence on the BFI-2 can be effectively addressed by within-person centering the item responses (thereby controlling acquiescence at the item level), by aggregating the items into their content-balanced domain and facet scales (thereby controlling acquiescence at the scale level), or by modeling acquiescence as a method factor (thereby accounting for the influence of acquiescence on item responses).

Convergence with other Big Five measures.

Domain-level associations. Our fourth objective for Study 3 was to examine the BFI-2's domain-level and facet-level associations with other Big Five measures. Table 9 presents domain-level correlations of the BFI-2 with the original BFI, as well as the BFAS, Mini-Markers, NEO-FFI, and NEO PI-R. These correlations show that the BFI-2 converged strongly with each of these measures. Specifically, the BFI-2's monotrait-heteromethod convergent correlations averaged .92 with the original BFI,[3] .82 with the BFAS, .80 with the Mini-Markers, .75 with the NEO-FFI, and .72 with the NEO PI-R. These correlations were much stronger than the absolute heterotrait-heteromethod discriminant correlations (M = .21), as well as the absolute heterotrait-monomethod discriminant correlations reported in Table 2 (M = .20 in the Internet sample and .24 in the student sample). Moreover, the BFAS, Mini-Markers, NEO-FFI, and NEO PI-R tended to converge more strongly with the BFI-2 than with the original BFI. This gain in convergence was most substantial for the Agreeableness (M = .75 for the BFI-2 vs. .70 for the BFI) and Open-Mindedness (M = .76 vs. .72) domains. These results indicate that broadening the Big Five domains' conceptual definitions from the BFI to the BFI-2 produced better domain-level convergence with other measures.[4]

Facet-level associations. To examine convergence and discrimination at the facet level, we computed each BFI-2 facet scale's correlations with the same-domain BFAS aspects and NEO PI-R facets. Figure 3 presents each BFI-2 facet's profile of correlations. This figure shows strong overall convergence: a total of 53 correlations were at least .50 in strength. Moreover, each BFI-2 facet showed a distinctive profile of associations with the BFAS and NEO PI-R. For example, within the Extraversion domain (Figure 3a), BFI-2 Assertiveness was much more strongly associ-

[3] Researchers interested in switching from the original BFI to the BFI-2 in ongoing longitudinal studies and research programs can obtain, from the first author, regression-derived scoring equations that further increase the comparability of scores on the BFI-2 with the original BFI. However, for new studies we recommend using simple item aggregation to score the BFI-2.

[4] This conclusion was further supported by the results of IRT analyses. These analyses indicated that, for each Big Five domain, the original BFI and the BFI-2 had similarly shaped test information curves but the BFI-2 provided more information. For example, compared with the original BFI, the BFI-2 provided approximately 55% more total information about Negative Emotionality and 15% more information about Extraversion; for both measures, information was distributed symmetrically between low and high levels of these two domains. Similarly, the BFI-2 provided approximately 29% more information about Agreeableness, 12% more information about Open-Mindedness, and 11% more information about Conscientiousness; both measures provided more information at very low levels than at very high levels of these three domains.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

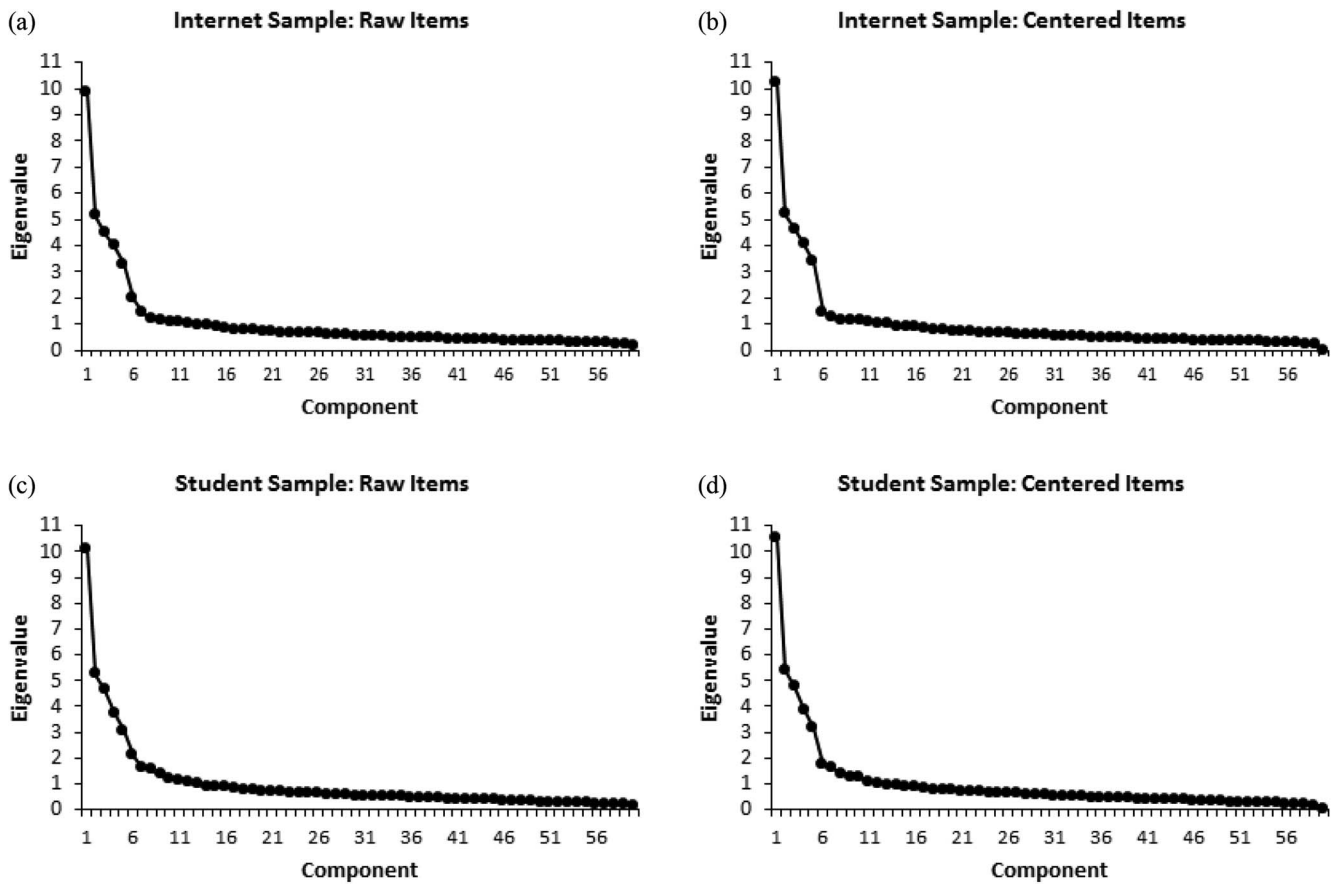

Figure 1. Eigenvalues from principal components analyses of the raw and centered BFI-2 items in the Internet and student validation samples (Study 3).

ated with BFAS Assertiveness (r = .85) and NEO PI-R Assertiveness (r = .79) than were BFI-2 Sociability and Energy Level. In contrast, BFI-2 Sociability was distinctively associated with Gregariousness (r = .62), whereas BFI-2 Energy Level was distinctively associated with Enthusiasm (r = .70), Positive Emotions (r = .63), and Activity (r = .59). Within the Agreeableness domain (Figure 3b), BFI-2 Compassion was distinctively associated with Compassion (r = .70), Altruism (r = .67), and Tender-Mindedness (r = .50), BFI-2 Trust with Trust (r = .65), and BFI-2 Respectfulness with Politeness (r = .66), Compliance (r = .52), and Straightforwardness (r = .48). Within Conscientiousness (Figure 3c), the most distinctive associations were BFI-2 Organization with Order (r = .74) and Orderliness (r = .70), BFI-2 Responsibility with Dutifulness (r = .58) and Deliberation (r = .47), and BFI-2 Productiveness with Industriousness (r = .77), Self-Discipline (r = .73), and Achievement-Striving (r = .60). Within Negative Emotionality (Figure 3d), the most distinctive associations were BFI-2 Depression with Withdrawal (r = .77) and NEO PI-R Depression (r = .71), BFI-2 Anxiety with NEO PI-R Anxiety (r = .70), and BFI-2 Emotional Volatility with Volatility (r = .79) and Angry Hostility (r = .54). Finally, for Open-Mindedness (Figure 3e), BFI-2 Aesthetic Sensitivity was most distinctively associated with Openness (r = .76) and Aesthetics (r = .74), BFI-2 Intellectual Curiosity with Ideas (r = .72) and Intellect (r = .62), and BFI-2 Creative Imagination with Fantasy (r = .38), especially the Fantasy item "I have a very active imagination" (r = .46).[5] All of these correlations were highly significant ($p < .001$), as were the corresponding partial correlations that controlled for same-domain BFI-2 facets. These results support the BFI-2 facets' conceptual definitions (see Table 1), as well as their convergent and discriminant validity.

Nomological network and predictive power. Our fifth and final objective for Study 3 was to examine the BFI-2's nomological network: its substantive relations with and capacity to predict conceptually relevant criteria (Cronbach & Meehl, 1955). We were particularly interested in testing whether the BFI-2's balance between bandwidth and fidelity would provide greater predictive power than the original BFI. We pursued this objective using a broad set of self-reported and peer-reported criteria. Table 10 presents each criterion's correlations with the BFI-2 domain scales, as well as standardized coefficients from regressions that

[5] The moderate degree of convergence between BFI-2 Creative Imagination and NEO PI-R Fantasy reflects Creative Imagination's greater focus on creativity and originality, as compared with Fantasy's greater focus on idle daydreaming.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 6
Loadings From a Principal Components Analysis of the 60 Within-Person Centered BFI-2 Items (Study 3)

Item (Original, revised, or new BFI item)	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality	Open-Mindedness
Extraversion					
Sociability items					
Tends to be quiet (Original)	**−.77/−.78**	.03/.06	.10/.09	−.01/−.03	.05/−.01
Is talkative (Original)	**.74/.79**	.11/.06	−.11/−.04	−.02/−.03	−.03/.07
Is outgoing, sociable (Original)	**.76/.73**	.24/.14	−.02/.04	.15/.18	−.03/.12
Is sometimes shy, introverted (Revised)	**−.71/−.71**	.07/.04	−.01/.08	−.16/−.20	.08/.09
Assertiveness items					
Is dominant, acts as a leader (New)	**.63/.65**	−.15/−.17	.26/.21	.13/.12	.19/.17
Has an assertive personality (Original)	**.58/.58**	−.27/−.29	.20/.15	.13/.13	.08/.06
Prefers to have others take charge (New)	**−.54/−.50**	.22/.22	−.18/−.23	−.11/−.09	−.23/−.21
Finds it hard to influence people (New)	**−.50/−.47**	.06/−.02	−.14/−.20	−.06/−.09	−.29/−**.35**
Energy Level items					
Is full of energy (Original)	**.67/.66**	.19/.24	.12/.08	.23/.18	.03/−.12
Shows a lot of Enthusiasm (Revised)	**.65/.68**	**.35/.31**	.08/.07	.05/.09	.15/.03
Rarely feels excited or eager (New)	**−.55/−.45**	−.29/−.29	.00/−.14	.05/.17	−.13/−.14
Is less active than other people (New)	**−.45/−.46**	−.06/.01	−.25/−.21	−.16/−.22	−.08/−.18
Agreeableness					
Compassion items					
Is compassionate, has a soft heart (New)	.07/.03	**.71/.67**	.06/.15	−.13/−.19	.11/.18
Can be cold and uncaring (Revised)	−.16/−.15	**−.65/−.69**	−.13/−.06	.04/.00	.01/−.07
Is helpful and unselfish with others (Original)	.18/.07	**.58/.61**	.19/.20	.05/.02	.14/.18
Feels little sympathy for others (New)	.01/−.05	**−.44/−.46**	−.03/−.13	.08/.12	−.13/−.23
Respectfulness items					
Is respectful, treats others with respect (New)	−.02/−.07	**.65/.61**	.19/**.33**	.08/.09	.11/.18
Is polite, courteous to others (New)	−.01/−.13	**.60/.62**	.19/.29	.05/.07	.10/.20
Is sometimes rude to others (Original)	.10/.23	**−.65/−.55**	−.17/−.09	−.13/−.17	−.01/−.07
Starts arguments with others (Revised)	.26/**.34**	**−.50/−.47**	−.17/−.06	−.18/−.18	.03/−.01
Trust items					
Assumes the best about people (New)	.10/.15	**.60/.63**	−.02/−.03	.15/.23	.03/.00
Has a forgiving nature (Original)	−.01/.02	**.61/.60**	.00/−.07	.24/.22	.05/.07
Tends to find fault with others (Original)	−.04/−.04	**−.52/−.60**	−.03/.05	−.28/−.20	−.04/−.02
Is suspicious of others' intentions (New)	−.17/−.17	**−.39/−.51**	.05/−.05	−.28/−.28	.06/.12
Conscientiousness					
Organization items					
Tends to be disorganized (Original)	−.03/.00	.01/.01	**−.75/−.76**	−.03/−.07	.06/.03
Is systematic, likes to keep things in order (New)	−.08/−.03	.02/−.01	**.70/.73**	.00/−.05	−.03/−.01
Keeps things neat and tidy (New)	.01/−.02	.00/.07	**.73/.69**	.00/.01	−.11/−.03
Leaves a mess, doesn't clean up (New)	−.03/.08	−.01/−.1	**−.69/−.62**	−.04/−.02	.07/−.05
Productiveness items					
Is efficient, gets things done (Revised)	.16/.25	.09/.01	**.68/.60**	.13/.15	.06/.11
Is persistent, works until the task is finished (Revised)	.08/.11	.13/.07	**.66/.60**	.13/.12	.06/.11
Tends to be lazy (Original)	−.23/−**.34**	−.10/−.10	**−.67/−.55**	−.07/−.09	−.05/−.06
Has difficulty getting started on tasks (New)	−.17/−.13	−.03/.02	**−.60/−.54**	−.12/−.16	.00/.14
Responsibility items					
Can be somewhat careless (Original)	.01/−.14	−.15/−.07	**−.61/−.55**	−.06/−.04	.03/−.14
Sometimes behaves irresponsibly (New)	.07/.02	−.17/−.15	**−.59/−.52**	−.10/−.09	.07/.08
Is reliable, can always be counted on (Revised)	.12/.09	**.30/.32**	**.54/.51**	.14/.13	.02/.18
Is dependable, steady (New)	.08/.09	.22/.18	**.51/.52**	.19/.14	.01/.15
Negative Emotionality					
Anxiety items					
Is relaxed, handles stress well (Original)	.05/.07	.04/.05	.10/.08	**.76/.76**	.07/.11
Worries a lot (Original)	−.18/−.19	.01/.04	.09/.08	**−.69/−.71**	.01/−.11
Rarely feels anxious or afraid (New)	.13/.12	−.01/−.09	.03/−.06	**.66/.65**	−.04/−.02
Can be tense (Original)	−.06/−.05	−.15/−.13	.04/.12	**−.59/−.59**	.05/−.01
Depression items					
Often feels sad (New)	**−.36/−.36**	−.09/−.15	−.17/−.19	**−.64/−.57**	.04/.03
Tends to feel depressed, blue (Revised)	**−.39/−.34**	−.08/−.19	−.14/−.25	**−.64/−.56**	.09/.09
Feels secure, comfortable with self (New)	**.33/.31**	.09/.14	.19/**.30**	**.58/.53**	.04/.02
Stays optimistic after experiencing a setback (New)	**.33/.31**	.15/.26	.20/.22	**.56/.46**	.12/.08
Emotional Volatility items					
Is emotionally stable, not easily upset (Original)	.00/.01	.09/.13	.13/.18	**.80/.77**	.05/.01
Is temperamental, gets emotional easily (New)	.10/.05	−.05/−.11	−.11/−.13	**−.73/−.71**	−.04/−.06
Keeps their emotions under control (New)	−.16/−.12	.05/.12	.23/.26	**.70/.64**	.07/−.02
Is moody, has up and down mood swings (Revised)	−.02/−.09	−.22/−.16	−.17/−.28	**−.64/−.64**	.03/.00

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 6 (*continued*)

Item (Original, revised, or new BFI item)	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality	Open-Mindedness
Open-Mindedness					
Intellectual Curiosity items					
Has little interest in abstract Ideas (New)	−.02/−.01	−.06/−.02	.09/−.06	−.07/−.02	**−.69/−.68**
Is complex, a deep thinker (Revised)	−.09/.02	−.03/−.02	.12/.18	−.12/−.02	**.56/.57**
Avoids intellectual, philosophical discussions (New)	−.06/−.15	.08/.01	.05/−.14	−.01/−.08	**−.54/−.52**
Is curious about many different things (Revised)	.04/.17	.15/**.30**	−.06/.13	−.05/−.02	**.50/.50**
Aesthetic Sensitivity items					
Is fascinated by art, music, or literature (Revised)	−.12/−.11	.09/.17	−.11/−.06	−.07/−.09	**.67/.69**
Has few artistic interests (Original)	−.03/.01	−.08/−.13	.05/.06	.04/.03	**−.64/−.63**
Values art and beauty (Revised)	−.03/−.11	.07/.17	.00/−.05	−.03/−.07	**.61/.64**
Thinks poetry and plays are boring (New)	−.01/.01	−.11/−.13	.01/−.02	.04/.18	**−.54/−.64**
Creative Imagination items					
Has little creativity (New)	−.15/−.17	−.07/−.02	.03/.05	.00/−.12	**−.66/−.68**
Is inventive, finds clever ways to do things (Revised)	.25/.09	−.06/−.03	.14/.04	.13/.22	**.55/.60**
Is original, comes up with new Ideas (Original)	.24/.18	.03/.05	.08/.02	.19/.27	**.57/.57**
Has difficulty imagining things (New)	−.12/−.12	−.10/−.11	.05/.01	−.10/.00	**−.54/−.51**

Note. BFI-2 items copyright 2016 by Oliver P. John and Christopher J. Soto. Reprinted with permission. Original = Item kept from the original BFI. Revised = Item revised from the original BFI. New = New item written for the BFI-2. Table entries are loadings on varimax-rotated principal components. Values left of each forward slash are for the Internet validation sample (N = 1,000); values right of each forward slash are for the student validation sample (N = 470). Absolute loadings of .30 or stronger are bolded. All items were within-person centered before analysis.

included all five domains as predictors.[6] Table 10 also presents the correlations and standardized coefficients for each criterion's strongest facet-level predictors; these facets were identified by forward regressions with an inclusion criterion of $p < .01$ and a maximum of three predictors.

The results shown in Table 10 reveal two important patterns. First, supporting the BFI-2's predictive breadth, each domain and facet scale significantly predicted at least one criterion; conversely, each criterion was significantly predicted by at least one domain and facet scale. Second, supporting the BFI-2's predictive specificity, each criterion was associated with a distinctive and conceptually meaningful set of predictors. Some criteria were clearly linked with a particular Big Five domain. For example, benevolent behavior was most strongly predicted by Agreeableness (especially Compassion and Trust), hedonistic behavior by low Conscientiousness (especially low Productiveness and Responsibility), self-directed behavior by Open-Mindedness (especially Intellectual Curiosity and Creative Imagination), peer-reported likability by Agreeableness (especially Trust), and peer-reported stress resistance by low Negative Emotionality (especially low Emotional Volatility and Anxiety). Other criteria were predicted by a combination of traits that spanned multiple domains. For example, power-seeking behavior was predicted by both low Agreeableness (especially Compassion and Respectfulness) and high Extraversion (especially Assertiveness), and personal growth well-being was predicted by a combination of Open-Mindedness (especially Intellectual Curiosity), Agreeableness (especially Compassion), and Extraversion (especially Energy Level).

To compare the predictive power of the BFI-2 with the original BFI, we conducted a series of regressions predicting each of the 20 self-reported and peer-reported criteria from either (a) the five BFI domain scales, (b) the 10 post hoc BFI facet scales, (c) the five BFI-2 domain scales, or (d) the 15 BFI-2 facet scales. The proportion of criterion variance explained by each regression is presented in Table 11, and the average predictive power for each set of predictors (i.e., mean R^2, averaged across all 20 criteria) is shown in Figure 4. This figure's leftmost bar shows that the original BFI domain scales collectively explained about one quarter of the criterion variance, on average (mean R^2 = .25). This represents a substantial degree of predictive power, but also leaves room for improvement. The next two bars illustrate that both the greater fidelity of the post hoc BFI facet scales (mean R^2 = .27) and the greater bandwidth of the BFI-2 domain scales (mean R^2 = .27) provided modest boosts to predictive accuracy over the original BFI domains. However, the most striking feature of Figure 4 is its rightmost bar, which shows that the BFI-2 facet scales provided considerably more predictive power (mean R^2 = .33) than any other set of predictors. Table 11 shows that this general pattern—with the BFI-2 facets clearly explaining the greatest proportion of criterion variance—generalized across the self-reported and peer-reported criteria.[7]

Taken together, these results support three key conclusions about the BFI-2's nomological network and predictive power.

[6] Each Big Five domain's pattern of associations with the self-reported and peer-reported criteria was very similar between the original BFI and the BFI-2. Column vector correlations comparing the two measures' patterns of zero-order correlations across the 20 criteria averaged .98, as did column vector correlations comparing the measures' patterns of standardized regression coefficients. However, associations with the criteria tended to be slightly stronger for the BFI-2 than for the original BFI, yielding the somewhat larger proportions of explained variance presented in Table 11 and Figure 4.

[7] This pattern was also consistent with hierarchical regression analyses conducted to test incremental changes in explained variance. For example, when the original BFI domain scales were entered as a first block of predictors, entering the post hoc BFI facet scales as a second block produced a significant increase in explained variance for 8 of the 20 criteria, entering the BFI-2 domains produced a significant increase for 10 criteria, and entering the BFI-2 facets produced a significant increase for 14 criteria. Similarly, when either the post hoc BFI facets or the BFI-2 domains were entered as a first block of predictors, entering the BFI-2 facets as a second block produced a significant increase for 14 of the 20 criteria.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 7
Loadings From a Principal Components Analysis of the 15 BFI-2 Facets (Study 3)

Facet	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality	Open-Mindedness
Extraversion					
Sociability	**.88/.89**	.06/.00	−.07/−.05	−.08/−.10	−.05/.02
Energy Level	**.78/.79**	**.26/.25**	.15/.17	−.12/−.12	.12/.07
Assertiveness	**.74/.72**	**−.22/−.22**	**.23/.24**	−.13/−.14	**.24/.24**
Agreeableness					
Compassion	.16/.13	**.83/.82**	.14/.17	.10/.11	.12/**.21**
Respectfulness	−.14/−**.27**	**.80/.74**	**.24/.24**	−.15/−.18	.05/.15
Trust	.12/.15	**.75/.79**	−.03/−.03	**−.32/−.30**	.02/.00
Conscientiousness					
Organization	.01/−.03	.00/.05	**.85/.83**	−.04/−.03	−.08/.00
Productiveness	**.22/.29**	.11/.04	**.83/.78**	−.13/−.16	.05/.03
Responsibility	.03/.09	**.26/.22**	**.78/.78**	−.16/−.14	−.03/.12
Negative Emotionality					
Anxiety	−.14/−.14	−.06/−.01	.00/.05	**.89/.88**	.01/−.07
Emotional Volatility	.05/.00	−.14/−.17	**−.20/−.25**	**.86/.83**	−.04/−.01
Depression	**−.43/−.41**	−.12/−**.20**	**−.21/−.28**	**.73/.67**	−.01/.01
Open-Mindedness					
Intellectual Curiosity	.01/.13	.01/.08	−.04/.16	.02/−.02	**.83/.78**
Aesthetic Sensitivity	−.03/−.07	.11/.18	−.05/−.04	.05/.12	**.80/.79**
Creative Imagination	**.26**/.18	.05/.03	.04/.01	−.12/−**.20**	**.76/.78**
Congruence	.99	.99	.99	1.00	.98

Note. Table entries are loadings on varimax-rotated principal components. Values left of each forward slash are for the Internet validation sample (N = 1,000); values right of each forward slash are for the student validation sample (N = 470). Absolute loadings of .20 or stronger are bolded. Congruence = Congruence coefficients comparing pairs of corresponding components between the Internet and student samples.

First, the BFI-2 domain and facet scales relate meaningfully with a variety of criteria, and these substantive relations support the scales' construct validity. Second, the BFI-2's hierarchical structure enhances its predictive power. The BFI-2 facet scales were developed to capture important distinctions within each Big Five domain, and these distinctions substantially increase the measure's capacity to predict self-reported and peer-reported criteria. Third, the BFI-2 provides greater predictive power than the original BFI at both the domain level and—especially—the facet level.

General Discussion

We developed the BFI-2 with four major goals in mind. To what extent did the present research achieve these goals? Our first goal was to develop a robust hierarchical structure. We addressed this goal by conceptualizing and measuring 15 traits that we propose constitute a minimally necessary set for capturing facet-level personality information within each Big Five domain (see Table 1; cf. DeYoung et al., 2007; John et al., 2008). We found that, at the domain level, the Big Five can be clearly recovered from both the BFI-2 items and facets (see Tables 6 and 7). At the facet level, the items within each domain can be modeled by three substantive facet factors plus an acquiescence method factor (see Table 8). This hierarchical structure, with three facets nested within each Big Five domain, is an important advance over the original BFI. Although the post hoc BFI facet scales offer some degree of differentiation within each domain, they do not constitute a true hierarchical structure (Nye, Allemand, Gosling, Potter, & Roberts, 2015). The BFI-2 thus extends the BFI's robust Big Five structure to the facet level.

Our second goal was to balance bandwidth and fidelity, and analyses of the BFI-2's nomological network indicate that it provides greater conceptual breadth and specificity, and thus greater predictive power, than the original BFI. Regarding bandwidth, the hierarchical model used to develop the BFI-2 ensured that each domain was broad enough to subsume three facet traits. This bandwidth allowed the BFI-2 domains to more strongly converge with other Big Five measures, and to more accurately predict a variety of criteria, than the original BFI domains (see Table 9 and Figure 4). Regarding fidelity, the 15 BFI-2 facets provide an additional level of conceptual specificity within the superordinate domains. This fidelity allowed the BFI-2 facets to show distinctive relations with self-reported and peer-reported criteria, and substantially enhanced their predictive power over the BFI and BFI-2 domains, as well as the post hoc BFI facet scales (see Table 10 and Figure 4).

Our third goal was to minimize the influence of acquiescent responding at both the scale and item levels, and the BFI-2 achieves this goal through balanced content. Some scales on the original BFI—especially Open-Mindedness and several of the post hoc facet scales—have a substantial imbalance of true-keyed and false-keyed items, allowing acquiescence to distort their associations with other variables (Danner, Aichholzer, & Rammstedt, 2015). In contrast, each BFI-2 domain and facet scale includes an equal number of true-keyed and false-keyed items. Thus, individual differences in acquiescence are automatically controlled at the scale level. The BFI-2's balanced content also makes it easier to address acquiescence at the item level, because simple within-person centering around each individual's mean item response effectively controls for acquiescence (see Figure 1). In contrast, the original BFI's imbalanced content means that simple within-person centering removes meaningful personality information along with acquiescence, a problem that can only be avoided by means of specially constructed (and rarely used) acquiescence scales (e.g., Soto et al., 2008, p. 737). Thus, compared with the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 8
Fit Statistics For Confirmatory Factor Analyses of the BFI-2 Items (Study 3)

Model	χ^2	*df*	BIC	CFI	TLI	RMSEA
Extraversion						
Single domain	1085.28/555.58	54	35282/15429	.786/.781	.738/.732	.138/.141
Single domain plus acquiescence	1073.86/531.10	53	35277/15410	.788/.791	.736/.740	.139/.139
Positive and negative items	1063.79/525.48	53	35267/15405	.790/.793	.739/.743	.138/.138
Three facets	389.25/251.77	51	34607/15143	**.930/.912**	**.909**/.886	.081/.092
Three facets plus acquiescence	331.78/204.92	50	34556/15103	**.941/.932**	**.923/.911**	**.075**/.081
Agreeableness						
Single domain	648.71/364.73	54	33394/14695	.805/.800	.762/.755	.105/.111
Single domain plus acquiescence	430.08/285.94	53	33183/14622	.876/.850	.846/.813	.084/.097
Positive and negative items	433.11/291.03	53	33186/14628	.875/.847	.845/.809	.085/.098
Three facets	492.31/243.22	51	33259/14592	.855/.876	.813/.840	.093/.090
Three facets plus acquiescence	196.82/128.96	50	32970/14484	**.952/.949**	**.936/.933**	**.054/.058**
Conscientiousness						
Single domain	1023.12/412.69	54	34085/15208	.793/.793	.747/.747	.134/.119
Single domain plus acquiescence	939.87/357.20	53	34008/15159	.811/.824	.765/.781	.129/.111
Positive and negative items	982.00/380.14	53	34050/15182	.802/.811	.753/.765	.132/.115
Three facets	508.63/233.29	51	33591/15047	**.902**/.895	.874/.864	.095/.087
Three facets plus acquiescence	338.36/133.23	50	33428/14953	**.939/.952**	**.919/.937**	**.076/.060**
Negative Emotionality						
Single domain	1144.80/609.03	54	35166/15683	.805/.779	.762/.730	.142/.148
Single domain plus acquiescence	1043.76/579.50	53	35072/15660	.823/.790	.780/.739	.137/.145
Positive and negative items	1012.23/563.46	53	35040/15644	.829/.797	.787/.747	.135/.143
Three facets	520.96/252.43	51	34563/15345	**.916/.920**	.891/.896	.096/.092
Three facets plus acquiescence	328.11/180.76	50	34377/15279	**.950/.948**	**.934/.931**	**.075/.075**
Open-Mindedness						
Single domain	839.22/632.83	54	33713/15235	.758/.697	.704/.629	.121/.151
Single domain plus acquiescence	814.63/619.58	53	33695/15227	.765/.703	.708/.630	.120/.151
Positive and negative items	814.92/618.33	53	33696/15226	.765/.704	.708/.631	.120/.151
Three facets	361.14/235.50	51	33256/14856	**.904/.903**	.876/.875	**.078**/.088
Three facets plus acquiescence	277.90/168.78	50	33179/14795	**.930/.938**	**.907/.918**	**.068/.071**

Note. BIC = Bayesian information criterion; CFI = Comparative fit index; TLI = Tucker-Lewis index; RMSEA = Root mean square error of approximation. Values left of each forward slash are for the Internet validation sample (N = 1,000); values right of each forward slash are for the student validation sample (N = 470). CFI and TLI values ≥ .900, and RMSEA values ≤ .080, are bolded.

original BFI, the BFI-2 is more resilient to the influence of acquiescence at both the scale and item levels.

Our final goal was for the BFI-2 to retain the focus, clarity, and brevity of the original BFI. To promote focus, we ensured that each BFI-2 domain and facet scale's conceptual definition was reflected in its item content. To promote clarity, we revised a number of original BFI items to elaborate their meaning or replace difficult vocabulary words. To ensure brevity, we limited the BFI-2 to 60 total items, so that the inventory can be completed in 10 min or less. We therefore expect that the BFI-2 will prove to be as conceptually coherent, easy to understand, and efficient to use as the original BFI.

Broader Implications and Lessons Learned

The present research has important implications for future studies examining personality structure and psychological assessment. One such implication is the utility of factor-pure and complementary facets. A persistent problem in the development of hierarchical measures is that the selection of lower-order constructs can influence a measure's higher-order structure in undesirable ways (Costa & McCrae, 1995; Goldberg, 1993a). For example, imagine that a hypothetical Big Five measure includes several facet scales conceptually related to both low Agreeableness and high Negative Emotionality (e.g., anger, antagonism, defensiveness, intolerance, irritability, moodiness, sarcasm, selfishness, suspicion, vanity). In this case, domain-level factor analyses may reveal that the either Agreeableness or Negative Emotionality factor has been unexpectedly rotated toward this interstitial cluster of closely related facets (to capture their shared variance), while the other domain factor has been rotated away from it (to capture the variance of facets falling outside this cluster). The BFI-2 addresses this potential problem by using a factor-pure facet—identified by previous research as central to its own domain and independent from the other four domains (Hofstee et al., 1992)—to partially define each of the Big Five. The present results indicate that this strategy successfully anchored the five domains in multidimensional space (see Table 7). Importantly, however, the BFI-2 also uses complementary facets to broaden each domain's definition beyond its factor-pure facet, and these complementary facets proved to be both conceptually and empirically valuable. For example, we found Organization to be a remarkably factor-pure facet of Conscientiousness. However, it is arguably the least conceptually interesting BFI-2 Conscientiousness facet, and Productiveness and Responsibility more strongly predicted several criteria (see Table 10). We therefore propose that broad constructs like the Big Five domains are best conceptualized and measured using a combination of factor-pure facets (to empirically anchor each superordinate construct) and complementary facets (to add conceptual and empirical

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

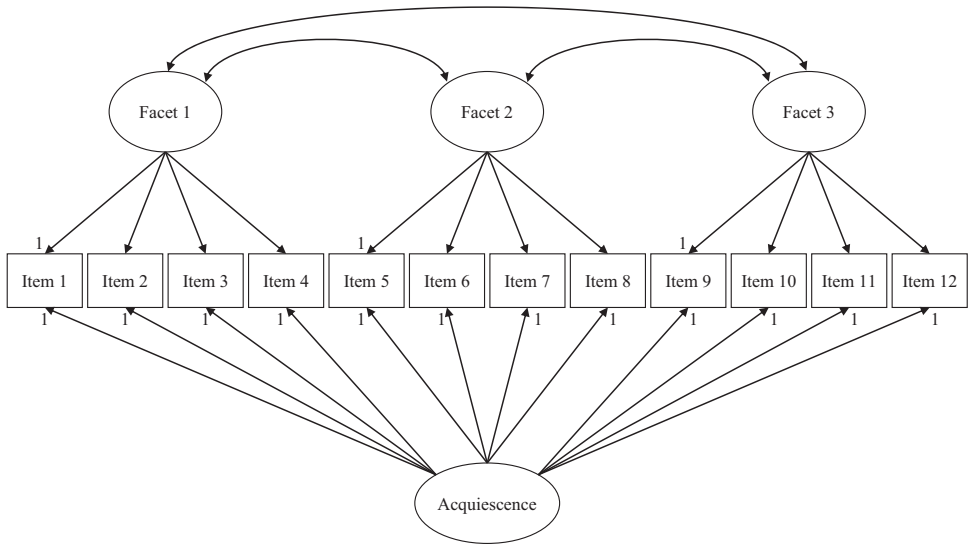

Figure 2. Illustration of the three facets plus acquiescence CFA model (Study 3).

breadth). Future research conducted to hierarchically measure sets of similarly broad constructs should benefit from this approach.

A second, related implication concerns the conceptualization of Open-Mindedness. Researchers have debated whether this domain should be primarily defined in terms of intellect, imagination, or openness to art and beauty (DeYoung, 2014). The present research addressed these alternative conceptualizations by including Intellectual Curiosity, Creative Imagination, and Aesthetic Sensitivity as the BFI-2's three Open-Mindedness facets. We found that all three facets were quite central to their domain (see Table 7). We therefore propose that Open-Mindedness is better conceptualized as broadly subsuming Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination than as narrowly focused on any one of them (cf. DeYoung et al., 2007).

A third implication concerns the bandwidth-fidelity tradeoff, especially the predictive power of broad (e.g., domains) versus narrow (e.g., facets) personality traits. Previous studies have found that narrow traits can provide greater predictive accuracy than broad traits across a variety of criteria (e.g., Ashton et al., 1995; Paunonen & Ashton, 2001). Similarly, the present findings indicate that the BFI and BFI-2 facet scales provide greater predictive power than their corresponding domain scales (see Figure 4). However, researchers have disagreed about whether the apparent incremental validity of narrow traits over broad ones reflects their greater fidelity or is simply a statistical artifact (cf. Ones & Viswesvaran, 1996; Paunonen, Rothstein, & Jackson, 1999). For example, compared with domain-level Big Five measures, facet-level measures provide a greater number of potential predictors; therefore, they may be more prone to capitalizing on chance in regressions predicting external criteria. The present findings argue against such an interpretation. If number of predictors were the primary determinant of predictive power, then we would expect a substantial boost in explained criterion variance from the five BFI and BFI-2 domains to the 10 post hoc BFI facet scales, followed by a smaller boost to the 15 BFI-2 facets (due to diminishing returns for each additional predictor). However, we found a very different pattern. The post hoc BFI facets provided only a modest increase in predictive power over the original BFI domains, and essentially no increase over the BFI-2 domains; the BFI-2 facets, however, provided a substantial increase in predictive power over the post hoc BFI facets (see Figure 4).

What explains this surprising pattern of results? We propose that it reflects differences in the approaches used to develop the BFI versus the BFI-2. The original BFI items were developed to maximize coherence within each Big Five domain, as well as differentiation between the domains (John, 1989, 1990; John & Srivastava, 1999). This focus on the domain level limited the ability of subsequent research to draw facet-level distinctions within each domain (Nye et al., 2015; Soto & John, 2009a). In contrast, development of the BFI-2 was guided, from the outset, by a hierarchical model of personality structure that drew clear distinctions both between and within the Big Five (see Table 1). This model facilitated the development of BFI-2 facet scales with greater fidelity (i.e., greater coherence within facets and better differentiation between facets) than their post hoc BFI counterparts. This greater fidelity, in turn, translated into greater predictive power. Thus, the present findings indicate that the predictive utility of narrow trait measures depends more on their quality than their quantity.

A fourth implication of the present research concerns acquiescence. Researchers have long known that acquiescent responding can bias the reliability, validity, and structure of psychological measures (Cronbach, 1946; Jackson & Messick, 1958). The present findings add to this literature by indicating that acquiescence does affect BFI-2 item responses, and therefore modestly biases the raw items' domain-level structure (see Figure 1) and more substantially biases their facet-level structure (see Table 8). Our findings also indicate that the BFI-2's content-balanced scales effectively control for acquiescence (see Table 7). Therefore, researchers interested in examining the BFI-2's scale-level associations with other variables can safely score each domain and facet scale using simple aggregation of the raw item responses. However, we strongly recommend that researchers interested in conducting item-level analyses of the BFI-2 account for the influence

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 9
Correlations of the BFI-2 and BFI Domains With the Mini-Markers, BFAS, NEO-FFI, and NEO PI-R in the Student Validation Sample (Study 3)

	BFI-2					BFI				
Domain	Ext.	Agr.	Con.	Neg.	Ope.	Ext.	Agr.	Con.	Neg.	Ope.
BFI										
Extraversion	**.93**	.02	.18	−.32	.10					
Agreeableness	.09	**.93**	.30	−.36	.23	.06				
Conscientiousness	.35	.27	**.91**	−.42	.20	.24	.27			
Negative Emotionality	−.31	−.29	−.29	**.94**	−.12	−.27	−.34	−.34		
Open-Mindedness	.22	.16	.07	−.11	**.87**	.14	.15	.14	−.13	
BFAS										
Extraversion	**.87**	.19	.34	−.40	.25	**.80**	.21	.43	−.35	.24
Agreeableness	−.02	**.73**	.23	−.09	.30	−.09	**.67**	.25	−.09	.17
Conscientiousness	.30	.15	**.82**	−.31	.05	.18	.14	**.80**	−.23	.00
Negative Emotionality	−.32	−.40	−.36	**.88**	−.17	−.28	−.45	−.42	**.86**	−.17
Open-Mindedness	.25	.22	.18	−.24	**.78**	.13	.21	.26	−.25	**.73**
Mini-markers										
Extraversion	**.88**	.08	.22	−.35	.14	**.90**	.13	.29	−.31	.17
Agreeableness	.11	**.80**	.31	−.22	.29	.06	**.77**	.31	−.19	.20
Conscientiousness	.26	.20	**.84**	−.34	.12	.14	.21	**.82**	−.28	.07
Negative Emotionality	−.15	−.46	−.35	**.74**	−.22	−.10	−.50	−.39	**.74**	−.19
Open-Mindedness	.22	.11	.16	−.14	**.75**	.12	.12	.22	−.17	**.75**
NEO-FFI										
Extraversion	**.71**	.29	.24	−.38	.18	**.69**	.34	.28	−.32	.21
Agreeableness	.01	**.74**	.23	−.26	.21	−.03	**.72**	.22	−.26	.14
Conscientiousness	.27	.21	**.80**	−.33	.16	.16	.21	**.79**	−.27	.14
Negative Emotionality	−.40	−.21	−.39	**.76**	−.11	−.37	−.25	−.44	**.74**	−.14
Open-Mindedness	.12	.21	.02	−.02	**.73**	.03	.18	.07	−.06	**.68**
NEO PI-R										
Extraversion	**.68**	.30	.14	−.22	.10	**.67**	.33	.19	−.17	.10
Agreeableness	−.25	**.71**	.05	−.05	.21	−.25	**.66**	.03	−.06	.12
Conscientiousness	.20	.08	**.74**	−.14	.18	.06	.06	**.74**	−.07	.14
Negative Emotionality	−.37	−.14	−.19	**.73**	−.01	−.38	−.21	−.25	**.73**	−.08
Open-Mindedness	.06	.25	−.01	−.02	**.74**	−.03	.21	.05	−.06	**.69**

Note. $N = 438$. Ext. = Extraversion; Agr. = Agreeableness; Con. = Conscientiousness; Neg. = Negative Emotionality; Ope. = Open-Mindedness. Convergent correlations are bolded. Absolute correlations of .10 or stronger are significant at $p < .05$.

of acquiescence by either centering each individual's set of item responses around their within-person mean or modeling acquiescence as a method factor (see Figure 2).

The BFI-2 can also be used in future research examining the phenomenon of acquiescent responding itself. For example, recent studies using the original BFI have shown that acquiescence effects tend to be especially pronounced among younger (Soto et al., 2008) and less well educated (Rammstedt & Farmer, 2013) respondents, as well as in more-collectivistic cultures (Rammstedt, Kemper, & Borg, 2013). However, this research has been limited to using subsets of BFI items, because of the measure's imbalanced content. In contrast, the BFI-2's fully content-balanced item set allows individual differences in acquiescence to be measured and clearly distinguished from meaningful personality information at multiple levels of abstraction: within a specific facet, within a Big Five domain, or across all five domains. For example, an individual's general acquiescence tendency can be indexed by simply averaging their 60 BFI-2 item responses without reversing the false-keyed items (e.g., an extreme yea-sayer who strongly agrees with all 60 items would receive an acquiescence score of 5.00, whereas an extreme naysayer who strongly disagrees with all 60 items would receive an acquiescence score of 1.00). Acquiescence can also be modeled as a latent variable that all items (both true-keyed and false-keyed) are constrained to load on equally (see Figure 2). Thus, the BFI-2 should prove useful for future research examining the causes, correlates, and consequences of acquiescence.

A final implication concerns the accuracy of personality self-reports and peer-reports. We found good overall self-peer agreement for the BFI-2, but also considerable variability in the degree of agreement across the Big Five domains and especially across their facets (see Table 4). Previous research has examined some possible determinants of self-peer agreement, such as evaluativeness and observability (e.g., John & Robins, 1993; Vazire, 2010). The BFI-2's hierarchical structure should prove useful for future work designed to further investigate this issue at the levels of the broad Big Five domains, narrower facets, and highly specific items.

Limitations and Future Directions

The present research had several important methodological strengths, including its large community, Internet, and student samples, its use of more than three thousand trait-descriptive adjectives and phrases as a rich source of personality-relevant item content, its integration of conceptual and empirical approaches to scale development, and its broad set of criteria for evaluating the BFI-2's basic measurement properties, multidimensional structure,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

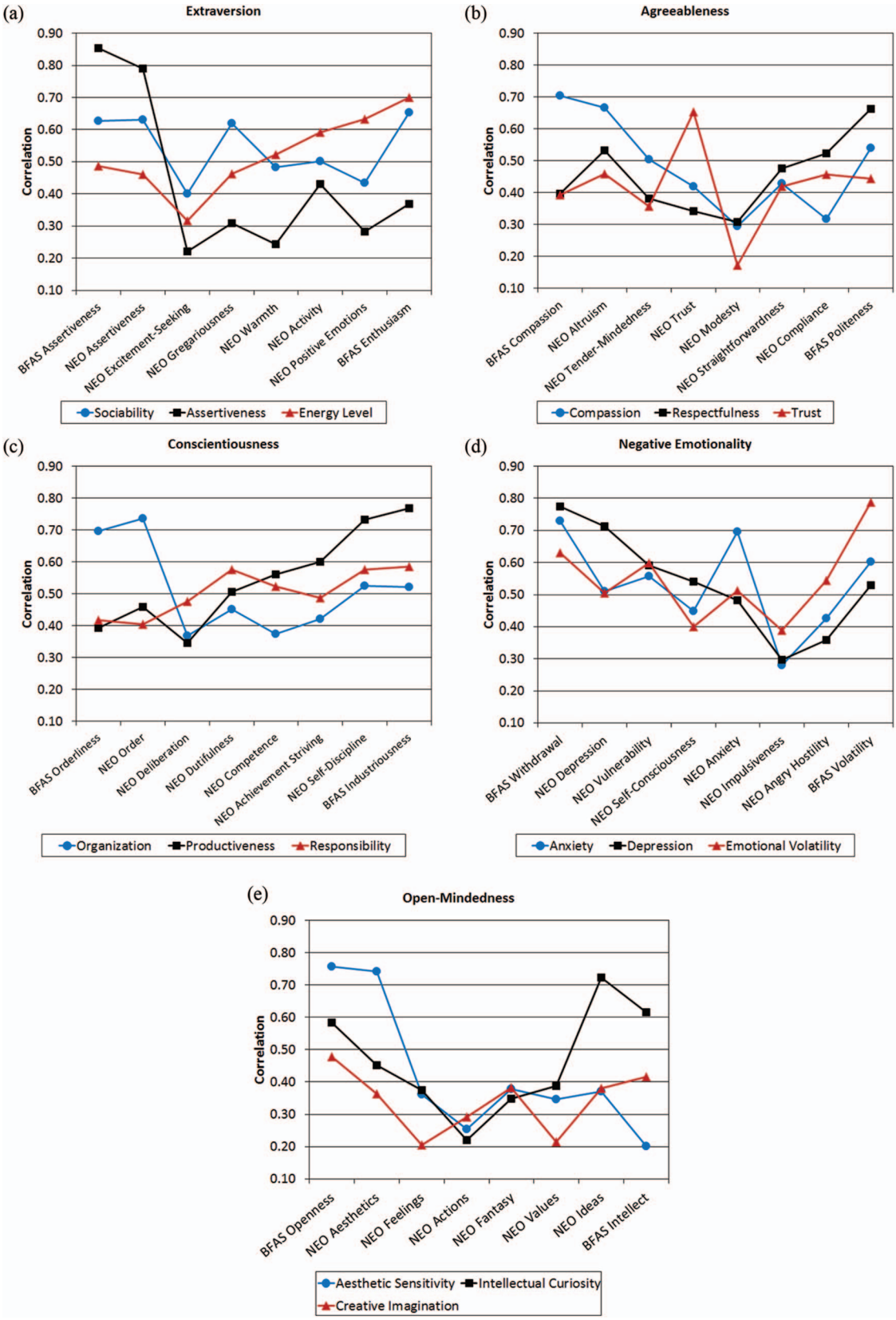

Figure 3. Correlations of the BFI-2 facets with the BFAS aspects and NEO PI-R facets (Study 3). $N = 438$. NEO = NEO PI-R. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 10
Associations of the BFI-2 Domains and Facets With Self-Reported and Peer-Reported Criteria (Study 3)

Criterion	Ext. *r*/β	Agr. *r*/β	Con. *r*/β	Neg. *r*/β	Ope. *r*/β	Strongest facet predictors (*r*/β)
Behavioral criteria						
Conformity	**−.35/−.32**	**.23/.31**	.08/.17	.17/.19	−.14/−.15	Respectfulness (.39/.38), Sociability (−.36/−.25), Creative Imagination (−.23/−.23)
Tradition	**−.24/−.27**	.02/.03	−.06/.03	−.05/−.04	−.14/−.16	Sociability (−.24/−.25), Intellectual Curiosity (−.21/−.19), Anxiety (−.07/−.14)
Benevolence	.11/.08	**.52/.52**	.18/.05	−.09/.12	**.23**/.10	Compassion (.47/.29), Trust (.45/.28), Intellectual Curiosity (.24/.14)
Power	**.30/.38**	**−.52/−.50**	−.11/−.07	.01/−.04	**−.21**/−.16	Assertiveness (.36/.31), Compassion (−.44/−.30), Respectfulness (−.51/−.29)
Universalism	.01/.00	**.20**/.19	−.02/−.08	−.04/.00	.15/.11	Trust (.21/.20), Aesthetic Sensitivity (.14/.13)
Hedonism	−.01/.10	**−.22**/−.12	**−.37/−.37**	.11/−.02	−.08/−.02	Productiveness (−.35/−.29), Responsibility (−.34/−.19), Sociability (.10/.18)
Security	−.18/**−.23**	.00/−.02	**.26/.37**	.03/.05	−.17/−.16	Organization (.30/.32), Sociability (−.19/−.18), Intellectual Curiosity (−.15/−.16)
Stimulation	.15/.19	−.16/−.15	**−.25/−.32**	−.03/−.10	.11/.15	Sociability (.21/.23), Responsibility (−.25/−.18), Organization (−.26/−.16)
Achievement	.15/.15	−.12/−.16	**.20/.25**	.01/.09	−.05/−.07	Productiveness (.26/.28), Trust (−.11/−.15)
Self-direction	.10/−.05	−.01/−.18	.07/.01	−.15/−.17	**.43/.47**	Intellectual Curiosity (.44/.39), Creative Imagination (.36/.19), Compassion (.00/−.13)
Psychological criteria						
Positive relations	**.47/.38**	**.46/.39**	**.39**/.09	**−.40**/−.13	**.24**/.03	Depression (−.53/−.38), Compassion (.41/.33), Sociability (.39/.25)
Purpose in life	**.47/.30**	**.30**/.16	**.54/.33**	**−.39**/−.12	**.24**/.06	Energy level (.53/.29), Depression (−.55/−.29), Responsibility (.49/.27)
Environmental mastery	**.49/.25**	**.29**/.09	**.57/.33**	**−.58/−.35**	.18/.00	Depression (−.65/−.51), Productiveness (.56/.35)
Self-acceptance	**.49/.28**	**.30**/.11	**.51/.23**	**−.57/−.35**	**.25**/.09	Depression (−.68/−.49), Responsibility (.47/.20), Energy level (.53/.20)
Autonomy	**.30**/.10	.19/−.02	**.36**/.19	**−.38/−.25**	**.35/.26**	Responsibility (.40/.30), Creative Imagination (.39/.26), Anxiety (−.32/−.25)
Personal growth	**.35/.24**	**.44/.31**	**.40**/.16	**−.28**/−.02	**.45/.28**	Intellectual Curiosity (.50/.37), Compassion (.48/.29), Energy level (.41/.26)
Peer-reported criteria						
Social connectedness	**.25/.20**	**.32/.27**	**.20**/.06	**−.27**/−.09	.04/−.03	Energy level (.33/.33), Respectfulness (.29/.28)
Likability	.16/.17	**.27/.28**	.10/.01	−.08/.06	.06/.01	Trust (.25/.25)
Stress resistance	**.24**/.09	.16/.01	.15/−.03	**−.51/−.49**	−.01/−.02	Emotional Volatility (−.47/−.32), Anxiety (−.46/−.28)
Positive affect	**.37/.26**	.09/−.02	.19/.05	**−.40/−.31**	−.04/−.07	Depression (−.42/−.35), Sociability (.32/.19)

Note. For behavioral criteria, $N = 439$ and correlations of .10 or stronger are significant at $p < .05$. For psychological criteria, $N = 179$ and correlations of .15 or stronger are significant at $p < .05$. For peer-reported criteria, $N = 184$ and correlations of .15 or stronger are significant at $p < .05$. Ext. = Extraversion; Agr. = Agreeableness; Con. = Conscientiousness; Neg. = Negative Emotionality; Ope. = Open-Mindedness; Positive relations = Positive relations with others. The domains' standardized coefficients are from multiple regressions that enter the five domains simultaneously. The strongest facet predictors were identified by forward regressions with an inclusion criterion of $p < .01$ and a maximum of three predictors. Domain correlations and standardized regression coefficients of .20 or stronger are bolded.

and nomological network. However, the present research also had some limitations that suggest important directions for future work. One such limitation is that most of the present participants were young or middle-aged adults with at least a high school level of education. Therefore, additional research using more representative samples will be needed to develop population norms for interpreting individual and group scores on the BFI-2. Moreover, previous research has shown that the original BFI can be completed by youths and adults with relatively little formal education, but that the measure tends to be somewhat less reliable and differentiated in these groups (Rammstedt & Farmer, 2013; Soto et al., 2008). While developing the BFI-2, we revised some BFI items to define or replace difficult vocabulary words. However, additional research will be needed to directly test whether these revisions have made the BFI-2 easier to understand, thereby leading to improved measurement properties among younger and less educated respondents. Research will also be needed to examine the BFI-2's reliability and validity among elderly adults.

A second, related limitation is that all of the present participants were residents of economically developed, English-speaking nations. The original BFI has now been translated into more than 25 languages and administered to members of more than 50 national cultures (e.g., Benet-Martínez & John, 1998; Denissen, Geenen, van Aken, Gosling, & Potter, 2008; Plaisant, Courtois, Réveillère, Mendelson, & John, 2010; Rammstedt & Farmer, 2013; Schmitt et al., 2007). Previous research has found both similarities and differences in the structure, mean levels, and correlates of the BFI across cultures, sparking debate about how these findings should be interpreted (e.g., Chan et al., 2012; Heine, Buchtel, & Norenzayan, 2008; Ramírez-Esparza, Mehl, Álvarez-Bermúdez, & Pennebaker, 2009). Future cultural research using the BFI-2 can help clarify these previous findings and investigate new questions. For example, we are curious to see whether and how the BFI-2's facet-level structure will vary across languages and cultures.

A third limitation concerns the brevity of the BFI-2. At 60 total items, the BFI-2 can efficiently measure the Big Five domains and

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 11
Predictive Power of the BFI and BFI-2 Domains and Facets (Study 3)

Criterion	BFI domains	BFI facets	BFI-2 domains	BFI-2 facets
Behavioral criteria				
Conformity	.27	.27	.25	.36
Tradition	.09	.12	.09	.14
Benevolence	.26	.28	.30	.32
Power	.37	.37	.42	.45
Universalism	.06	.06	.06	.09
Hedonism	.15	.16	.16	.22
Security	.15	.18	.17	.21
Stimulation	.15	.16	.15	.20
Achievement	.09	.11	.10	.14
Self-direction	.23	.24	.23	.27
Mean	.18	.20	.19	.24
Psychological criteria				
Positive relations	.46	.48	.45	.51
Purpose in life	.38	.41	.43	.51
Environmental mastery	.50	.52	.54	.58
Self-acceptance	.46	.49	.51	.58
Autonomy	.27	.32	.28	.38
Personal growth	.36	.38	.42	.49
Mean	.40	.43	.44	.51
Peer-reported criteria				
Social connectedness	.17	.20	.18	.27
Likability	.09	.08	.10	.17
Stress resistance	.29	.29	.27	.33
Positive affect	.23	.25	.23	.29
Mean	.19	.20	.19	.26
Overall mean	.25	.27	.27	.33

Note. For behavioral criteria, $N = 439$. For psychological criteria, $N = 179$. For peer-reported criteria, $N = 184$. Each table entry is the R^2 value for a criterion variable predicted from a set of domain or facet scales.

15 key facet traits, but cannot comprehensively measure all important personality characteristics. Thus, deciding what content to include on the BFI-2 necessarily entailed excluding some content. These inclusion versus exclusion decisions occurred at three levels. Some occurred at the domain level: because the BFI-2 is designed to measure the Big Five, it excludes some important personality traits, as well as attributes such as characteristic adaptations and narrative identity, that fall beyond the Big Five (e.g., McAdams & Pals, 2006; Paunonen & Jackson, 2000).

A second set of exclusions occurred at the facet level, when we selected three key facets within each Big Five domain. Our goal was to focus on facets that have been conceptually and empirically prominent in the personality literature. Thus, the BFI-2 includes facets that match closely with all 10 aspects of the BFAS, and with many of the 30 facets assessed by the NEO PI-R (see Figure 3). However, limiting the BFI-2 to only three facets per domain meant excluding some potential facets. For example, the BFI-2 focuses on cognitive and aesthetic, rather than behavioral, facets of Open-Mindedness. This decision reflects questionnaire-based and lexical research indicating that cognitive and aesthetic content tends to be most prominent within this domain (DeYoung et al., 2007; McCrae & Costa, 2008; Saucier & Ostendorf, 1999). However, willingness to try new behaviors certainly remains an important personality characteristic. Other potential facets that were excluded due to their relatively peripheral position within the Big Five domains, but that remain important traits in their own right, include excitement-seeking, modesty, deliberation, and impulsivity (see Figure 3).

A final set of exclusions occurred at the item level, when we selected adjectives and phrases to operationalize the BFI-2 facets. Specifically, establishing the BFI-2's hierarchical structure required us to exclude some interstitial content that fell between multiple domains or facets. For example, when developing the BFI-2 Emotional Volatility facet we initially considered item content that explicitly assessed both moodiness/instability and anger/irritability. However, analyses conducted during Study 1 indicated that anger/irritability content fell in the interstitial space between the Negative Emotionality and Agreeableness domains, and between the Emotional Volatility and Depression facets. Therefore, the final BFI-2 Emotional Volatility scale directly assesses moodiness/instability, and relies on the strong association between moodiness/instability and anger/irritability to indirectly measure the latter content (as shown by BFI-2 Emotional Volatility's strong convergence with two scales that include considerable anger/irritability content: BFAS Volatility and NEO PI-R Angry Hostility; see Figure 3d).

These exclusion decisions at the domain, facet, and item levels highlight that, although the BFI-2 is a highly efficient, hierarchically structured Big Five measure, its brevity means that it cannot be entirely comprehensive. Thus, there is also considerable value to longer measures that directly assess personality content falling between and beyond the Big Five.

Conclusion

Taken together, the present findings indicate that the BFI-2 is a reliable and valid measure of the Big Five domains and their key facets, whose nomological network includes substantive relations with a variety of self-reported and peer-reported criteria. Our findings further indicate that the BFI-2 represents an important advance over the original BFI. Specifically, the BFI-2 introduces a robust hierarchical structure, minimizes the influence of acquies-

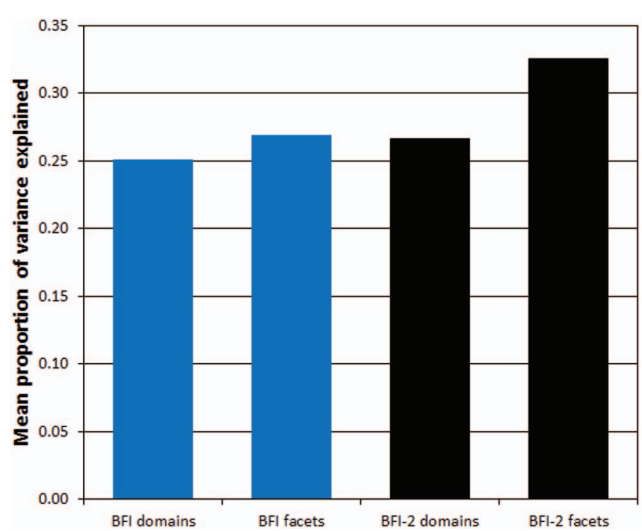

Figure 4. Mean R^2 values for regression analyses predicting 20 self-reported and peer-reported criteria from the BFI and BFI-2 domains and facets (Study 3). See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

cent responding, and provides greater bandwidth, fidelity, and predictive power than the BFI, while still retaining the original measure's conceptual focus, ease of understanding, and brevity. The BFI-2 therefore offers valuable new opportunities for research examining the structure, assessment, development, and life outcomes of personality traits, and we look forward to the results of this future work.

References

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. http://dx.doi.org/10.1006/jrpe.1995.1025

Bardi, A., & Schwartz, S. H. (2003). Values and behavior: Strength and structure of relations. *Personality and Social Psychology Bulletin, 29,* 1207–1220. http://dx.doi.org/10.1177/0146167203254602

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26. http://dx.doi.org/10.1111/j.1744-6570.1991.tb00688.x

Benet-Martínez, V., & John, O. P. (1998). *Los Cinco Grandes* across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75,* 729–750. http://dx.doi.org/10.1037/0022-3514.75.3.729

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

Chan, W., McCrae, R. R., De Fruyt, F., Jussim, L., Löckenhoff, C. E., De Bolle, M., . . . Terracciano, A. (2012). Stereotypes of age differences in personality traits: Universal and accurate? *Journal of Personality and Social Psychology, 103,* 1050–1066. http://dx.doi.org/10.1037/a0029712

Clark, L. A., & Watson, D. (2008). Temperament: An organizing paradigm for trait psychology. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 265–286). New York, NY: Guilford Press.

Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64,* 21–50. http://dx.doi.org/10.1207/s15327752jpa6401_2

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322–331. http://dx.doi.org/10.1037/0022-3514.81.2.322

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement, 6,* 475–494.

Cronbach, L. J., & Gleser, G. C. (1957). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. http://dx.doi.org/10.1037/h0040957

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality, 57,* 119–130. http://dx.doi.org/10.1016/j.jrp.2015.05.004

Denissen, J. J. A., Geenen, R., van Aken, M. A. G., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment, 90,* 152–157. http://dx.doi.org/10.1080/00223890701845229

DeYoung, C. G. (2014). Openness/Intellect: A dimension of personality reflecting cognitive exploration. In M. L. Cooper & R. J. Larsen (Eds.), *APA handbook of personality and social psychology: Personality processes and individual differences* (Vol. 4, pp. 369–399). Washington, DC: American Psychological Association.

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. http://dx.doi.org/10.1037/0022-3514.93.5.880

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92,* 1087–1101. http://dx.doi.org/10.1037/0022-3514.92.6.1087

Ekman, P. (1992). Are there basic emotions? *Psychological Review, 99,* 550–553. http://dx.doi.org/10.1037/0033-295X.99.3.550

English, T., John, O. P., Srivastava, S., & Gross, J. J. (2012). Emotion regulation and peer-rated social functioning: A 4-year longitudinal study. *Journal of Research in Personality, 46,* 780–784. http://dx.doi.org/10.1016/j.jrp.2012.09.006

Eysenck, H. (1967). *The biological basis of personality*. Springfield, IL: Thomas.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26–42. http://dx.doi.org/10.1037/1040-3590.4.1.26

Goldberg, L. R. (1993a). The structure of personality traits: Vertical and horizontal aspects. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 169–188). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10127-024

Goldberg, L. R. (1993b). The structure of phenotypic personality traits. *American Psychologist, 48,* 26–34. http://dx.doi.org/10.1037/0003-066X.48.1.26

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84–96. http://dx.doi.org/10.1016/j.jrp.2005.08.007

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. http://dx.doi.org/10.1037/0022-3514.48.1.82

Gough, H. G. (1957). *Manual for the California Psychological Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G., & Heilbrun, A. B., Jr. (1983). *The Adjective Check List manual*. Palo Alto, CA: Consulting Psychologists Press.

Graziano, W. G., & Tobin, R. M. (2009). Agreeableness. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 46–61). New York, NY: Guilford Press.

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science, 19,* 309–313. http://dx.doi.org/10.1111/j.1467-9280.2008.02085.x

Hendriks, A. A. J. (1997). *The construction of the Five-Factor Personality Inventory (FFPI)*. Groningen, The Netherlands: Rijksuniversiteit Groningen.

Hirsh, J. B., DeYoung, C. G., Xiaowen Xu., & Peterson, J. B. (2010). Compassionate liberals and polite conservatives: Associations of agreeableness with political ideology and moral values. *Personality and Social Psychology Bulletin, 36,* 655–664. http://dx.doi.org/10.1177/0146167210366854

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Hofstee, W. K., de Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63,* 146–163. http://dx.doi.org/10.1037/0022-3514.63.1.146

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

John, O. P. (1989). Towards a taxonomy of personality descriptors. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 261–271). New York, NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4684-0634-4_20

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York, NY: Guilford Press.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology, 60,* 348–361. http://dx.doi.org/10.1037/0022-3514.60.3.348

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 Press.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61,* 521–551. http://dx.doi.org/10.1111/j.1467-6494.1993.tb00781.x

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. 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 Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York, NY: Guilford Press.

Lucas, R. E., Le, K., & Dyrenforth, P. S. (2008). Explaining the extraversion/positive affect relation: Sociability cannot account for extraverts' greater happiness. *Journal of Personality, 76,* 385–414. http://dx.doi.org/10.1111/j.1467-6494.2008.00490.x

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences, 19,* 451–458. http://dx.doi.org/10.1016/j.lindif.2009.03.007

Markowitz, E. M., Goldberg, L. R., Ashton, M. C., & Lee, K. (2012). Profiling the "pro-environmental individual": A personality perspective. *Journal of Personality, 80,* 81–111. http://dx.doi.org/10.1111/j.1467-6494.2011.00721.x

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist, 61,* 204–217. http://dx.doi.org/10.1037/0003-066X.61.3.204

McCrae, R. R. (1994). Openness to Experience: Expanding the boundaries of Factor V. *European Journal of Personality, 8,* 251–272. http://dx.doi.org/10.1002/per.2410080404

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19,* 97–112. http://dx.doi.org/10.1177/1088868314541857

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 Press.

McCrae, R. R., Herbst, J. H., & Costa, P. T. (2001). Effects of acquiescence on personality factors structures. In R. Riemann, F. M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217–231). Berlin, Germany: Pabst Science Publishers.

Merriam-Webster. (2005). *Merriam-Webster's collegiate dictionary* (11th ed.). Springfield, MA: Author.

Nye, C. D., Allemand, M., Gosling, S. D., Potter, J., & Roberts, B. W. (2015). Personality trait differences between young and middle-aged adults: Measurement artifacts or actual trends? *Journal of Personality,* n/a. http://dx.doi.org/10.1111/jopy.12173

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17,* 609–626. http://dx.doi.org/10.1002/(SICI)1099-1379(199611)17:6<609::AID-JOB1828>3.0.CO;2-K

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57,* 401–421. http://dx.doi.org/10.1146/annurev.psych.57.102904.190127

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. http://dx.doi.org/10.1037/0022-3514.81.3.524

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the big five? Plenty! *Journal of Personality, 68,* 821–835. http://dx.doi.org/10.1111/1467-6494.00117

Paunonen, S. V., Rothstein, M. G., & Jackson, D. N. (1999). Narrow reasoning about the use of broad personality measures for personnel selection. *Journal of Organizational Behavior, 20,* 389–405. http://dx.doi.org/10.1002/(SICI)1099-1379(199905)20:3<389::AID-JOB917>3.0.CO;2-G

Plaisant, O., Courtois, R., Réveillère, C., Mendelsohn, G. A., & John, O. P. (2010). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse Convergente avec le NEO-PI-R [Factor structure and internal reliability of the French Big Five Inventory (BFI-Fr). Convergent and discriminant validation with the NEO-PI-R]. *Annales Médico-Psychologiques, 168,* 97–106. http://dx.doi.org/10.1016/j.amp.2009.09.003

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135,* 322–338. http://dx.doi.org/10.1037/a0014996

Pozzebon, J. A., & Ashton, M. C. (2009). Personality and values as predictors of self- and peer-reported behavior. *Journal of Individual Differences, 30,* 122–129. http://dx.doi.org/10.1027/1614-0001.30.3.122

Ramírez-Esparza, N., Mehl, M. R., Álvarez-Bermúdez, J., & Pennebaker, J. W. (2009). Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study. *Journal of Research in Personality, 43,* 1–7. http://dx.doi.org/10.1016/j.jrp.2008.09.002

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment, 25,* 1137–1145. http://dx.doi.org/10.1037/a0033323

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,* 71–81. http://dx.doi.org/10.1002/per.1894

Roberts, B. W., Bogg, T., Walton, K. E., Chernyshenko, O. S., & Stark, S. E. (2004). A lexical investigation of the lower-order structure of conscientiousness. *Journal of Research in Personality, 38,* 164–178. http://dx.doi.org/10.1016/S0092-6566(03)00065-5

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. http://dx.doi.org/10.1111/j.1744-6570.2005.00301.x

Rogers, C. R. (1954). Toward a theory of creativity. *Etc., 11,* 249–260.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social*

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Psychology, 57, 1069–1081. http://dx.doi.org/10.1037/0022-3514.57.6.1069

Saucier, G. (1992). Openness versus Intellect: Much ado about nothing. *European Journal of Personality, 6,* 381–391. http://dx.doi.org/10.1002/per.2410060506

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment, 63,* 506–516. http://dx.doi.org/10.1207/s15327752jpa6303_8

Saucier, G. (1997). Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology, 73,* 1296–1312. http://dx.doi.org/10.1037/0022-3514.73.6.1296

Saucier, G., & Goldberg, L. R. (2003). The structure of personality attributes. In M. R. Barrick & A. M. Ryan (Eds.), *Personality and work: Reconsidering the role of personality in organizations* (pp. 1–29). San Francisco, CA: Jossey-Bass.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology, 76,* 613–627. http://dx.doi.org/10.1037/0022-3514.76.4.613

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38,* 173–212. http://dx.doi.org/10.1177/0022022106297299

Sneed, C. D., McCrae, R. R., & Funder, D. C. (1998). Lay conceptions of the five-factor model and its indicators. *Personality and Social Psychology Bulletin, 24,* 115–126. http://dx.doi.org/10.1177/0146167298242001

Soto, C. J., & John, O. P. (2009a). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality, 43,* 84–90. http://dx.doi.org/10.1016/j.jrp.2008.10.002

Soto, C. J., & John, O. P. (2009b). Using the California Psychological Inventory to assess the Big Five personality domains: A hierarchical approach. *Journal of Research in Personality, 43,* 25–38. http://dx.doi.org/10.1016/j.jrp.2008.10.005

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,* 718–737. http://dx.doi.org/10.1037/0022-3514.94.4.718

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100,* 330–348. http://dx.doi.org/10.1037/a0021717

Ten Berge, J. M. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research, 34,* 89–102. http://dx.doi.org/10.1207/s15327906mbr3401_4

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T., Jr. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging, 20,* 493–506. http://dx.doi.org/10.1037/0882-7974.20.3.493

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98,* 281–300. http://dx.doi.org/10.1037/a0017908

Watson, D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 767–793). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-012134645-4/50030-5

Widiger, T. A. (2009). Neuroticism. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 129–146). New York, NY: Guilford Press.

(*Appendix follows*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix

The Big Five Inventory-2 Self-Report Form 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. __ Is outgoing, sociable
2. __ Is compassionate, has a soft heart
3. __ Tends to be disorganized
4. __ Is relaxed, handles stress well
5. __ Has few artistic interests
6. __ Has an assertive personality
7. __ Is respectful, treats others with respect
8. __ Tends to be lazy
9. __ Stays optimistic after experiencing a setback
10. __ Is curious about many different things
11. __ Rarely feels excited or eager
12. __ Tends to find fault with others
13. __ Is dependable, steady
14. __ Is moody, has up and down mood swings
15. __ Is inventive, finds clever ways to do things
16. __ Tends to be quiet
17. __ Feels little sympathy for others
18. __ Is systematic, likes to keep things in order
19. __ Can be tense
20. __ Is fascinated by art, music, or literature
21. __ Is dominant, acts as a leader
22. __ Starts arguments with others
23. __ Has difficulty getting started on tasks
24. __ Feels secure, comfortable with self
25. __ Avoids intellectual, philosophical discussions
26. __ Is less active than other people
27. __ Has a forgiving nature
28. __ Can be somewhat careless
29. __ Is emotionally stable, not easily upset
30. __ Has little creativity
31. __ Is sometimes shy, introverted
32. __ Is helpful and unselfish with others
33. __ Keeps things neat and tidy
34. __ Worries a lot
35. __ Values art and beauty
36. __ Finds it hard to influence people
37. __ Is sometimes rude to others
38. __ Is efficient, gets things done
39. __ Often feels sad
40. __ Is complex, a deep thinker
41. __ Is full of energy
42. __ Is suspicious of others' intentions
43. __ Is reliable, can always be counted on
44. __ Keeps their emotions under control
45. __ Has difficulty imagining things
46. __ Is talkative
47. __ Can be cold and uncaring
48. __ Leaves a mess, doesn't clean up
49. __ Rarely feels anxious or afraid
50. __ Thinks poetry and plays are boring
51. __ Prefers to have others take charge
52. __ Is polite, courteous to others
53. __ Is persistent, works until the task is finished
54. __ Tends to feel depressed, blue
55. __ Has little interest in abstract Ideas
56. __ Shows a lot of Enthusiasm
57. __ Assumes the best about people
58. __ Sometimes behaves irresponsibly
59. __ Is temperamental, gets emotional easily
60. __ Is original, comes up with new Ideas

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

Note. BFI-2 items copyright 2016 by Oliver P. John and Christopher J. Soto. Reprinted with permission.

(*Appendix continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Scoring Key

Item numbers for the BFI-2 domain and facet scales are presented below. False-keyed items are denoted by "R." SPSS syntax for scoring the scales is available from the first author. To download or get more information about the BFI-2, visit the Colby Personality Lab website (http://www.colby.edu/psych/personality-lab/).

BFI-2 Domain Scales

Extraversion: 1, 6, 11R, 16R, 21, 26R, 31R, 36R, 41, 46, 51R, 56

Agreeableness: 2, 7, 12R, 17R, 22R, 27, 32, 37R, 42R, 47R, 52, 57

Conscientiousness: 3R, 8R, 13, 18, 23R, 28R, 33, 38, 43, 48R, 53, 58R

Negative Emotionality: 4R, 9R, 14, 19, 24R, 29R, 34, 39, 44R, 49R, 54, 59

Open-Mindedness: 5R, 10, 15, 20, 25R, 30R, 35, 40, 45R, 50R, 55R, 60

BFI-2 Facet Scales

Sociability: 1, 16R, 31R, 46
Assertiveness: 6, 21, 36R, 51R
Energy Level: 11R, 26R, 41, 56
Compassion: 2, 17R, 32, 47R
Respectfulness: 7, 22R, 37R, 52
Trust: 12R, 27, 42R, 57
Organization: 3R, 18, 33, 48R
Productiveness: 8R, 23R, 38, 53
Responsibility: 13, 28R, 43, 58R
Anxiety: 4R, 19, 34, 49R
Depression: 9R, 24R, 39, 54
Emotional Volatility: 14, 29R, 44R, 59
Intellectual Curiosity: 10, 25R, 40, 55R
Aesthetic Sensitivity: 5R, 20, 35, 50R
Creative Imagination: 15, 30R, 45R, 60

Received September 4, 2015
Revision received February 19, 2016
Accepted February 19, 2016 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.