Development and validation of the Danish Big Five Inventory-2: Domain- and facet-level structure, construct validity, and reliability

Vedel, Anna; Wellnitz, Kaare Bro; Ludeke, Steven; Soto, Christopher J.; John, Oliver P.; Andersen, Simon Calmar

Published in:
European Journal of Psychological Assessment

DOI (link to publication from Publisher):
10.1027/1015-5759/a000570

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Development and Validation of the Danish Big Five Inventory-2: Domain- and Facet-level Structure, Construct Validity, and Reliability

Abstract

Following the publication of the Big Five Inventory-2 (BFI-2) and its abbreviated forms (the 30-item BFI-2-S and 15-item BFI-2-XS), two studies were conducted to develop and validate a Danish translation of these measures. Study 1 first developed a preliminary Danish BFI-2 item pool consisting of translations of the 60 BFI-2 items, then tested and refined this item pool using two waves of data collection, and identified a set of 60 item formulations for the Danish BFI-2. Study 1 then examined the domain- and facet-level structure of the Danish BFI-2, and the construct validity and reliability of this measure. Study 2 tested the generalizability of the measurement properties of the Danish BFI-2 found in Study 1 as well as the preliminary measurement properties of its abbreviated forms (the Danish BFI-2-S and BFI-2-XS) in a new sample. The results of these studies indicate that the Danish BFI-2 is a reliable and valid personality measure with psychometric properties and construct validity corresponding to the English-language original. The preliminary results regarding measurement properties of the abbreviated forms are encouraging and should inspire further validation.

Keywords: Big Five; Five-Factor Model; facets; personality measurement; test translation

1. Introduction

1.1. The Big Five: definitions, predictive validity, and cross-cultural utility

Individual differences in the way people think, feel, and behave can be organized in terms of the Big Five personality trait domains: Extraversion (reflecting the tendency to be sociable, assertive, and active); Agreeableness (compassionate, accommodating, and generous), Conscientiousness (orderly, diligent, and responsible), Negative Emotionality (anxious, ruminating, and moody; alternatively labeled Neuroticism or by its opposite pole, Emotional Stability), and Open-Mindedness (curiosity about diverse intellectual and cultural experiences; alternatively labeled Openness to Experience, Intellect, or Imagination; Goldberg, 1993; John, Naumann, & Soto, 2008; McCrae & Costa, 2008). Furthermore, these five broad personality domains can be conceptualized hierarchically, with each domain subsuming a number of narrower personality facets (DeYoung, Quilty, & Peterson, 2007; Goldberg, 1999; McCrae & Costa, 2010).

A large body of research has examined the validity of the Big Five personality domains, and the domains have been found to predict important life outcomes such as physical and mental health, divorce and occupational attainment (Kotov, Gamez, Schmidt, & Watson, 2010; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007; Steel, Schmidt, & Shultz, 2008). With the growing consensus around the Big Five domains as a useful organizing personality framework, the robustness of the Big Five has also been explored in many different cultures, with results generally supporting the cross-cultural utility of the five domains (Benet-Martínez & Oishi, 2008; McCrae et al., 2005; Saucier & Goldberg, 2001). Recently, an improved instrument for efficiently measuring the five personality domains, as well as three facets within each of them, was developed by Soto and John (2017a): The Big Five Inventory-2 (BFI-2). The present study was conducted

to develop and validate a translation of the BFI-2, and to test whether the same trait and facet structure could be found and replicated in two Danish samples.

1.2. Development of the BFI-2 and its abbreviated forms

The BFI-2 (Soto & John, 2017a) is a 60-item measure of the Big Five domains and 15 narrower facets: Extraversion (with facets of Sociability, Assertiveness, and Energy Level), Agreeableness (Compassion, Respectfulness, and Trust), Conscientiousness (Organization, Productiveness, and Responsibility), Negative Emotionality (Anxiety, Depression, and Emotional Volatility), and Open-Mindedness (Intellectual Curiosity, Aesthetic Sensitivity, and Creative Imagination). The BFI-2 is a major revision of the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991), an inventory measuring the Big Five domains with 44 short and easily comprehensible phrases. The BFI-2 constitutes an important advance. Unlike the BFI, the BFI-2 was deliberately developed to offer a robust hierarchical structure with three facets nested within each Big Five domain. The greater conceptual breadth (at the domain level) and specificity (at the facet level) provided in the BFI-2 due to this hierarchical structure leads to greater predictive power. Also, the BFI-2 domain and facet scales were constructed with an equal number of true-keyed and false-keyed items, effectively controlling for acquiescent response style (the tendency of an individual to consistently agree or consistently disagree with items, regardless of their content; Jackson & Messick, 1958), which was not the case with the BFI. Finally, the BFI-2 was developed with the aim to retain the focus and brevity of the BFI despite the inclusion of explicitly measured facets: with a total of 60 items in the BFI-2, the inventory can be completed in less than 10 minutes (Soto & John, 2017a).

Despite the relative brevity of the BFI-2 compared with many other Big Five measures, there are situations in which an even shorter measure is needed. In large-scale

surveys, for example, where many different variables are to be measured, it may not be possible to devote more than a few minutes to the assessment of personality traits. Also, in research where participants are asked to rate themselves multiple times or asked to rate themselves as well as others, a very brief measure might be needed in order to prevent participant fatigue and careless responding. Concerns about participant fatigue and exhaustion may also make very brief measures the preferred choice when working with particular populations, such as children or the elderly. Consequently, Soto and John (2017b) developed a 30-item short form (the BFI-2-S) and a 15-item extra-short form (BFI-2-XS) of the BFI-2. At the domain level, these abbreviated forms retain much of the reliability and validity of the full measure. At the facet level, the abbreviated forms fare less well, though the BFI-2-S is still useful for the assessment of facets in reasonably large samples (Soto & John, 2017b).

1.3. Overview of the present research

The present research was conducted to develop and validate a Danish translation of the BFI-2 and its abbreviated forms, the BFI-2-S and the BFI-2-XS. These translations would allow efficient, hierarchical assessment of the Big Five personality domains in the Danish cultural context and cross-cultural research. We pursued this goal through two studies. Study 1 first developed a preliminary Danish BFI-2 item pool consisting of translations of the BFI-2 items, then tested and refined this item pool using two waves of data collection, and identified a set of 60 item formulations for the final Danish BFI-2. Study 1 then examined the domain- and facet-level structure of the Danish BFI-2, as well as the construct validity and reliability of this measure. Study 2 tested the generalizability of the measurement properties of the Danish BFI-2 found in Study 1, as well as the preliminary measurement properties of its abbreviated forms (the Danish BFI-2-S and BFI-2-XS) in a new sample.

2. Study 1

Study 1 had two main goals. The first was to develop a preliminary pool of Danish BFI-2 item translations, test and refine this pool, and identify a final set of 60 item formulations for the Danish BFI-2. The second was to examine the preliminary domain- and facet-level structure of the Danish BFI-2, its construct validity, and its reliability. To assess preliminary construct validity, we administered an alternative measure of the Big Five, as well as measures of affect and psychological well-being. We expected high convergent correlations between the Danish BFI-2 scales and their respective counterparts in the alternative Big Five measure, and based on previous research using other Big Five measures (Steel et al., 2008), we also expected Negative Emotionality to correlate positively with negative affect and negatively with psychological well-being and expected Extraversion to correlate positively with positive affect and psychological well-being.

2.1. Method

2.1.1. Participants

Study 1 analyzed data from two samples collected at two time points, T1 (February 2017) and T2 (April 2017). The sample in which the first wave of data was collected (T1 sample) consisted of first-year political science students (N = 65; M_{age} = 21.6 (SD = 1.5); 64.6% female) and second-year psychology students (N = 137; M_{age} = 23.5 (SD = 4.6); 83.2% female) from a Danish university. The sample in which the second wave of data was collected (T2 sample) consisted of a subset of 101 participants from the T1 sample (M_{age} = 22.8, SD = 4.1 at T1; 81.2% female).

2.1.2. Procedure

2.1.2.1. Data collection. All data were collected online using generic survey links, and participants completed the surveys either at a lecture or at home. Students who participated in both waves of data collection entered a lottery to win movie tickets. In the first wave of data collection (T1), participants rated themselves using the preliminary pool of 87 Danish BFI-2 translations (for translation procedures, see https://osf.io/rmgv4/). Participants also completed three additional measures to establish convergent and discriminant validity (see Measures section). The participants in the T1 sample were invited to provide their email address so that they could be contacted for a test-retest follow-up study two months later. 179 participants provided their email address, and 101 of these completed the follow-up survey in the second wave of data collection (T2). The follow-up survey was limited to the corrected pool of translations (79 items; see the following section and supplementary material provided at https://osf.io/rmgv4/).

2.1.2.2. Development of the first and second BFI-2 item pools. The first pool of 87 candidate items for the Danish BFI-2 was developed using a translation, back-translation, and alternative-translation procedure described in the supplementary material available from https://osf.io/rmgv4/. After administering this pool to the T1 sample, reliability analysis and exploratory principal component analysis (PCA; see https://osf.io/rmgv4/) were conducted to select the best-fuctioning item translations. One could argue that exploratory factor analysis (EFA) is a more appropriate approach, given the general conception of measured personality characteristics as reflections of latent (personality) factors. For this reason all further analyses after the initial item selection in Study 1, and all analyses in Study 2, were EFAs. As the present study aimed

to closely model the validation of the English-language BFI-2, in which PCA was employed, PCAs were also conducted, though, and can be found in the supplementary material at https://osf.io/rmgv4/. These yielded results similar to the EFAs. For most BFI-2 items, the preliminary item pool included a translation that worked well, but a few formulations seemed redundant, and six BFI-2 items required new Danish formulations (see https://osf.io/rmgv4/). Based on these initial analyses and reformulations, a corrected pool of 79 BFI-2 item translations was administered to the T2 sample.

2.1.3. Measures

2.1.3.1. Big Five Inventory-2 (BFI-2). The BFI-2 (Soto & John, 2017a) is a hierarchical measure of the Big Five personality domains and 15 narrower facet traits. The measure's 60 items are short, descriptive phrases with the common item stem "I am someone who...", followed by item-specific content (e.g., "Is outgoing, sociable"). Respondents rate themselves on each item using a 5-point Likert scale ranging from *disagree strongly* to *agree strongly*. Soto and John (2017a) provided evidence for the structure, reliability, and validity of the BFI-2 domain and facet scales. Alpha reliabilities of the 12-item domain scales averaged .87 in their study, with a total range of .83 to .90 across samples. Alpha reliabilities of the four-item facet scales averaged .76 and .77 in two validation samples, with a total range of .66 to .85 across samples. Alphas in the present study are presented in Tables 1 and 2, and discussed below.

2.1.3.2. Positive and Negative Affect Schedule (PANAS). The PANAS (Watson, Clark, & Tellegen, 1988) consists of two 10-item mood scales measuring positive and negative affect. Each item presents an affective state (e.g., "excited" or "distressed"), and respondents rate the extent to which they have experienced this state during the last

week on a 5-point scale ranging from *very slightly or not at all* to *very much*. The PANAS was translated into Danish (see https://osf.io/rmgv4/ for translation procedures), and alpha reliability of both the Positive Affect (PA) and Negative Affect (NA) scales was .83.

2.1.3.3. Psychological Well-Being scales (PWB). The PWB developed by Ryff (1989) measures different aspects of healthy psychological functioning in adults. The measure comprises six scales: Self-acceptance, Positive relations with others, Autonomy, Environmental mastery, Purpose in life, and Personal growth. Several versions of the PWB are in use (Abbott, Ploubidis, Huppert, Kuh, & Croudace, 2010). In the present study, a Danish short version of the PWB (Jønsson, 2016) was used, including 18 items (three items per scale). All items were short phrases with positive item content (e.g., "In general, I feel confident and positive about myself"), and respondents rated themselves on a 6-point Likert scale ranging from *disagree strongly* to *agree strongly*. Given the limited number of items for each well-being scale, only the total PWB scores were used. Alpha reliability of this overall well-being scale was .88.

2.1.3.4. Big Five Mini-Markers. The Mini-Markers (Saucier, 1994) are a 40-item short form of Goldberg's (1992) unipolar Big Five marker adjectives, with 8 items measuring each Big Five domain. All items are trait-descriptive adjectives that respondents rate on a 9-point Likert scale ranging from *extremely inaccurate* to *extremely accurate* as a description of the respondent. Mini-Markers were translated into Danish (see https://osf.io/rmgv4/ for translation procedures), and alpha reliabilities for the Mini-Marker scales were .85 for Extraversion, .84 for Agreeableness, .91 for Conscientiousness, .81 for Negative Emotionality, and .80 for Open-Mindedness.

2.1.4. Statistical analyses

First, 60 item formulations were selected for the Danish BFI-2 based on 1) an exploratory PCA (see https://osf.io/rmgv4/) including all 79 items administered at T2, and 2) facet-level alpha reliabilities for different item constellations (see https://osf.io/rmgv4/). We then subjected these items to further analyses to provide preliminary tests of reliability and validity. To begin these analyses, random intercept EFA with varimax rotation was done using the 60 selected items to test the domain-level structure of the Danish BFI-2. A second EFA was then conducted using the 15 facet mean scores. Next, test-retest reliability was calculated using data from the 101 students who participated in both waves of data collection. Retest reliabilities were estimated based on item sets with the greatest possible overlap from the first to second wave of data collection. Lastly, convergent and discriminant validity were examined through correlational analyses using the T1 dataset. Because the results of these analyses may be biased by the fact that we used the same data set for item selection and scale validation, we present the results of Study 1 briefly.

2.2. Results

2.2.1. Reliability and intercorrelations

Tables 1 and 2 present alpha reliability coefficients and interscale correlations for the BFI-2 domain scales and facet scales, respectively, in the T2 sample (see https://osf.io/rmgv4/ for descriptive statistics). The alpha reliabilities and interscale correlations were calculated based on the 60 items selected for further analyses. As shown in Table 1, all domain scales had alpha reliabilities above .85 in the T2 sample. As shown in Table 2, alpha reliabilities of all but two facets were above .70, which appears reasonable considering that each facet is measured by only four items (of which

two are reverse-coded).

The retest reliabilities of the domains ranged between .85 and .92 (see Table 1), exceeding the values obtained by Soto and John (2017a). The retest reliabilities of the facets ranged between .66 and .92 (see Table 2) and were also generally higher than the corresponding values of Soto and John (2017a). Because only six of the 60 initial item formulations differed (most of them very little) from the final chosen formulations, the test-retest analysis provides reasonable estimates of the test-retest stability of the Danish BFI-2.

Scale intercorrelations supported the domain and facet structure of the Danish BFI-2. Absolute correlations between the domain scales averaged only .20 (see Table 1). Regarding the facet scales, within-domain facet correlations averaged .59, whereas absolute between-domain facet correlations averaged .17 (see Table 2).

2.2.2. Big Five factor structure. The domain-level structure of the Danish BFI-2 was tested using a random intercept EFA on the 60 chosen items in the T2 sample, effectively controlling for acquiescence. All scree plots, outputs from parallel analyses[1], and loadings are presented in the supplementary material available at https://osf.io/rmgv4/.

As expected, the scree plot as well as parallel analysis suggested the extraction of five factors. The EFA produced a clear Big Five structure, with all items loading most strongly, and at least .39 in magnitude, on their intended domain component. To further examine the domain-level structure of the Danish BFI-2, a EFA of the 15 facet scales was conducted using the mean facet scores. The scree plot and the parallel

[1] All parallel analyses were conducted using the syntax from O'Connor (2000).

analyses both suggested the extraction of five factors. In the five-factor solution all facets loaded .57 or above on the intended domain and had negligible secondary loadings, in line with the results from Soto and John (2017a).

2.2.3. Construct validity

Table 3 presents correlations of the Danish BFI-2 domains with the Mini-Markers, the PANAS, and the PWB based on the T1 dataset. As expected, all five personality domains as measured by the BFI-2 correlated highly with the same domain as measured by the Mini-Markers. Specifically, the BFI-2's monotrait-heteromethod convergent correlations averaged .80 with the Mini-Markers, whereas the absolute heterotrait-heteromethod discriminant correlations averaged only .18. Consistent with previous research, BFI-2 Negative Emotionality correlated .60 with PANAS Negative Affect and -.55 with Psychological Well-being, whereas Extraversion correlated .40 with Positive Affect and .58 with Psychological Well-Being. These results provide encouraging preliminary evidence regarding the construct validity of the Danish BFI-2.

3. Study 2

Two key limitations of Study 1 were that (a) the selected set of 60 Danish BFI-2 item formulations differed somewhat between T1 and T2, and (b) the same data sets were used for scale development and validation, potentially leading to biased estimates of measurement properties. Therefore, Study 2 tested the generalizability of the measurement properties of the Danish BFI-2 found in Study 1 as well as the preliminary measurement properties of its abbreviated forms (the Danish BFI-2-S and BFI-2-XS) in a new sample.

3.1. Method

3.1.1. Participants and procedure

Study 2 analyzed data from 287 adults collected in June and July 2017. All data were collected online using survey links distributed through different digital channels for Danish psychology students. Participants completed the survey at home and entered a lottery to win movie tickets in return. A simple attentiveness question was part of the survey, and only data from participants answering this question correctly were included in the analyses (19 out of the initial 306 participants were excluded based on the attentiveness question). Participants' age ranged from 19 to 53, with most being in their mid-twenties (M_{age} = 25.75, *SD* = 5.85). The majority were female (79.6%), and most were psychology students (62.3%) or students of another subject (27.1%). Smaller proportions of participants were working (6.7%) or currently unemployed (3.9%).

3.1.2. Measures

Participants rated themselves using the Danish BFI-2 items developed and refined in Study 1. Two items (item 7 and 26) were tested with two different translations, since these items had two formulations each that fared equally well in the initial analyses conducted in Study 1, resulting in a total item pool of 62 Danish BFI-2 items to be administered. Based on loadings from a new EFA (see Statistical analyses), the best formulation of item 7 and 26 was retained, and the two inferior versions were discarded, leading to a final set of 60 Danish BFI-2 items used for all analyses reported below (at the time of publication available from https://osf.io/rmgv4/).

3.1.3. Statistical analyses

The statistical analyses largely replicated the approach of Study 1. EFAs with varimax

rotation were conducted using (a) the final 60 items, and (b) the 15 facet scales, to test the domain-level structure of the final Danish BFI-2. To test the facet-level structure within each domain, the 60 items were also subjected to confirmatory factor analyses (CFA). In addition to the analyses pertaining to the full BFI-2, additional EFAs using (a) the 30 items included in the BFI-2-S, and (b) the 15 items included in the BFI-2-XS, were conducted to explore the domain-level structure of the Danish abbreviated forms (these analyses and results should be interpreted with caution, though, see section 3.2.3 and https://osf.io/rmgv4/). As in Study 1, all EFA results were replicated using PCA.

3.2. Results

3.2.1. Reliability and intercorrelations for the BFI-2

Tables 1 and 2 present reliability coefficients and interscale correlations for the BFI-2 domain scales and facet scales, respectively (see https://osf.io/rmgv4/ for descriptive statistics). As shown in Table 1, all domain scales had alpha reliabilities above .81, and as shown in Table 2, alpha reliabilities of all but four facets were above .70, which appears reasonable considering that each facet is measured by only four items. The overall pattern of reliability coefficients was very similar to Study 1, as indicated by a column-vector correlation of .87.

Scale intercorrelations again supported the domain and facet structure of the Danish BFI-2 (see Table 1). Absolute correlations between the domain scales averaged only .15. The overall pattern of domain scale intercorrelations was very similar to Study 1 and to Soto and John (2017a), as indicated by column-vector correlations of .93 and

.85, respectively.[2] Regarding the facet scales, within-domain facet correlations averaged .48, whereas absolute between-domain correlations averaged .13 (see Table 2). Column-vector correlations comparing the pattern of facet intercorrelations to Study 1 and to Soto and John (2017a), were .91 and .89, respectively.

3.2.2. Multidimensional structure of the BFI-2

All scree plots, parallel analyses outputs, EFA loadings, and PCA replication loadings are available at https://osf.io/rmgv4/.

3.2.2.1. BFI-2 domain-level structure. The scree plot suggested the extraction of five or six factors, and the parallel analysis suggested the extraction of six factors. We therefore examined both five- and six-factor solutions. The five-factor solution revealed a clear Big Five structure, with all items loading most strongly on their intended domain. Each factor consisted of 12 items with primary loadings mostly above .50 and mostly negligible secondary loadings. Only item 11 had somewhat ambiguous loadings, with similar loadings on both its intended Extraversion component and the Agreeableness component. Congruence coefficients comparing each Study 2 factor with the corresponding Study 1 factor ranged from .92 to .95[3]. The six-factor solution was

---

[2] All column-vector correlations in this section comparing results from the Danish BFI-2 to results from Soto and John (2017a) are based on Soto and John's student validation sample, which seems most comparable to the present sample.

[3] Congruence coefficients were computed using the R package "psych" based on the formula, $\frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$, where X and Y are column vectors of component loadings from the two compared samples, and XY connotes element wise multiplication, i.e., x_1y_1, x_2y_2, etc. (see also ESM).

similar, but split the Extraversion factor into two: one comprising the Assertiveness facet and one comprising the Sociability and Energy Level facets.

To further examine the domain-level structure of the Danish BFI-2, a EFA of the 15 facet scales was conducted using the mean facet scores. The scree plot as well as the parallel analyses suggested the extraction of five factors. All facets loaded .50 or above on the intended domain and had negligible secondary loadings. Congruence coefficients comparing the Study 2 loadings with those from Study 1 ranged from .96 to .98. Taken together, the results replicated the intended domain-level structure of the BFI-2.

3.2.2.2. BFI-2 facet-level structure. Following the statistical procedures by Soto and John (2017a), we tested multidimensional structure at the facet level using confirmatory factor analysis (CFA). Specifically, a series of five CFA models were fit to the raw items within each Big Five domain. Fit statistics for these models are presented in Table 4. The first *single domain* model allowed all 12 items of a given domain to load on a single factor. As expected, this model showed a poor fit for all domains, thereby suggesting that more complex models should be tested.

The next two models allowed for individual differences in acquiescence. The *single domain plus acquiescence* model allowed all 12 items of a given domain to load on both a domain factor and an acquiescence factor. All loadings on the acquiescence factor were constrained to equal 1, and the acquiescence factor was not allowed to correlate with the domain factor. These constraints ensured that differences in response style would be distinct from the personality content of the domain factor (see Soto & John, 2017a). The *positive and negative items* model split true-keyed and false-keyed items into two separate factors that were allowed to correlate. As expected, these two

models yielded essentially identical fit, and fit remained unacceptable for all five domains.[4]

The fourth and fifth model both tested the three-dimensional facet structure expected to underlie each of the BFI-2 domains. The *three facets* model included three factors representing the three facet scales within each Big Five domain. Each item was only allowed to load on a single facet factor, and the three facet factors were allowed to intercorrelate. This model provided a substantial improvement in fit for all domains compared to the three previously tested models. The final *three facets plus acquiescence* model added an acquiescence factor (see https://osf.io/rmgv4/ for a graphical display). In this model, each item was allowed to load on both its facet factor as well as an acquiescence factor, which was not allowed to correlate with any of the facet factors. This model provided acceptable overall fit for each domain. Consistent with the results of Soto and John (2017a), these results indicate that the facet-level structure of the BFI-2 within each Big Five domain can be adequately modelled by three facet factors and an acquiescence method factor.

3.2.3. Preliminary results for the BFI-2-S and BFI-2-XS

In addition to the analyses of the full BFI-2, we examined the preliminary measurement properties of the BFI-2-S and BFI-2-XS abbreviated forms. As the full BFI-2 and the abbreviated forms were not administered independently, we describe these results only briefly, refer the reader to the supplementary material available at https://osf.io/rmgv4/ for further preliminary results, and caution that these results should be interpreted cautiously pending independent replication (see Smith, McCarthy, & Anderson, 2000).

[4] Convergence of these models for the Open-Mindedness domain required an alternative specification in which the factor variances were constrained to equal one.

Overall, the Danish short forms showed preliminary measurement properties comparable to the English-language short forms. Alpha reliabilities of the BFI-2-S domain scales averaged .74, and alphas of the BFI-2-XS domain scales averaged .58. The preliminary analyses produced a clear Big Five structure for both the BFI-2-S and the BFI-2-XS, with all items but one (in the BFI-2-XS only) loading most strongly on their intended domain. Moreover, congruence coefficients comparing the Study 2 loadings with the results of Soto and John (2017b) were all .93 or above, except for a congruence coefficient of .46 for Open-Mindedness in the BFI-2-XS caused by *stronger* primary loadings for two Danish items.

4. General discussion

The present research was conducted to develop and validate a Danish translation of the BFI-2 and its abbreviated forms, the BFI-2-S and the BFI-2-XS. First, we developed a preliminary pool of candidate item translations, then identified a final set of 60 item formulations for the Danish BFI-2, and examined the multidimensional structure and measurement properties of this measure. Second, we tested the generalizability of the multidimensional structure and the measurement properties of the Danish BFI-2, as well as of its abbreviated forms, in a new sample. The domain- and facet-level structure of the Danish BFI-2 replicated the BFI-2's intended hierarchical structure, with three facets nested within each Big Five domain. The measurement properties of the Danish BFI-2 also corresponded to the measurement properties of the English-language BFI-2 (Soto & John, 2017a), as did the domain- and facet-level intercorrelations. Furthermore, associations of the Danish BFI-2 with the PANAS, Mini-Markers, and Psychological Well-Being provide preliminary support for the construct validity of the Danish BFI-2. The preliminary results regarding the measurement properties of the abbreviated measures, the Danish BFI-2-S and BFI-2-XS, were also consistent with the English-

language measures, though further, independent validation of these abbreviated measures is necessary.

4.1. Limitations and future directions

The present research had several important strengths, including its translation process involving both language and content experts, its refinement of a preliminary item pool into the final Danish BFI-2, and its consideration of multiple reliability and validity criteria. However, this research also had some limitations that highlight directions for future research. First, the samples employed in the studies were medium-sized, and replication of the results in larger samples would be desirable. The samples also consisted of students primarily (in the case of Study 1, students exclusively). Students are a highly select group (Henrich, Heine, & Norenzayan, 2010), and those in different academic majors also tend to have different personality profiles (Vedel, 2016). Since the majority of our participants were psychology students, and psychology students as a group tend to have high levels of Negative Emotionality and Open-Mindedness (Vedel, 2016), the mean values in the present studies, and particularly of these traits, will most likely be different from the mean values in the general population. Mean values of the Big Five traits in the present studies therefore should not be regarded as norms, even for students. Furthermore, future studies can examine the generalizability of the results for the Danish BFI-2 and its abbreviated forms in more diverse samples.

A second limitation pertains to the revision of items across the two studies. Revising some item formulations to achieve better measurement properties and clearer multidimensional structure was deemed necessary. However, since the retest-analyses and the construct validity analyses were conducted using the data from Study 1, this meant that these analyses were conducted on items where a few had slightly different formulations than the final Danish BFI-2 items. Future studies should therefore examine

the generalizability of the test-retest reliability and the construct validity found in the present studies. Future studies might also compare the properties of the Danish BFI-2 with other Danish Big Five measures, such as the Danish IPIP-NEO-120 (Vedel, Gøtzsche-Astrup, & Holm, 2018).

A third limitation is that the results regarding the measurement properties of the Danish BFI-2-S and BFI-2-XS should be interpreted with caution. As noted above, these short measures and the full Danish BFI-2 were not administered independently, which may affect their measurement properties (Smith, McCarthy, & Anderson, 2000). Further validation of these abbreviated measures is therefore warranted. Finally, we echo the call of Soto and John (2017b) for large-scale studies with high statistical power testing the utility of the abbreviated forms in terms of facet-level measurement.

5. Conclusion

Taken together, the present findings indicate that the Danish BFI-2 is a reliable and valid measure of the Big Five domains and their subsumed facets. The high degree of similarity between the measurement properties of the Danish BFI-2, as compared with the English-language BFI-2, supports the validity of the Danish translation. Furthermore, the multidimensional structure of the Danish measures replicates the BFI-2's intended Big Five and facet-level structure. The present research also suggest that the Danish BFI-2-S and BFI-2-XS have measurement properties similar to their English-language counterparts, though further validation is needed to confirm these preliminary results. Translations such as the current one will hopefully expand the BFI-2 into a wide range of countries and cultures, thereby providing a shared conceptual and measurement framework for advancing cross-cultural personality research.

Acknowledgments

This project was financially supported by TrygFonden's Centre for Child Research.

References

Abbott, R. A., Ploubidis, G. B., Huppert, F. A., Kuh, D., & Croudace, T. J. (2010). An Evaluation of the Precision of Measurement of Ryff's Psychological Well-Being Scales in a Population Sample. *Social Indicators Research, 97*(3), 357-373. https://doi.org/10.1007/s11205-009-9506-x

Benet-Martínez, V., & Oishi, S. (2008). Culture and personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research, Third Edition* (pp. 542-567). New York, NY: Guilford.

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*(5), 880-896. https://doi.org/10.1037/0022-3514.93.5.880

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*(1), 26-42. https://doi.org/10.1037/1040-3590.4.1.26

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34. https://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. J. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (pp. 7-28). Tilburg, NL: Tilburg University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences, 33*(2-3), 111-135. https://doi.org/10.1017/S0140525X0999152X

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

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

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.

Jønsson, T. F. (2016). *The Danish version of the Psychological Well-being scales*. Unpublished manuscript.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin, 136*(5), 768-821. https://doi.org/10.1037/a0020327

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,Third Edition* (pp. 159-181). New York, NY: Guilford.

McCrae, R. R., & Costa, P. T. (2010). *NEO Inventories professional manual*. Lutz, FL: Psychological Assessment Resources.

McCrae, R. R., Terracciano, A., & Members of the Personality Profiles of Cultures Project. (2005). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology, 88*(3), 547-561. https://doi.org/10.1037/0022-3514.88.3.547

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc, 32*(3), 396-402. https://doi.org/10.3758/bf03200807

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science, 2*(4), 313-345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Ryff, C. D. (1989). Happiness Is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being. *Journal of Personality and Social Psychology, 57*(6), 1069-1081. https://doi.org/10.1037/0022-3514.57.6.1069

Saucier, G. (1994). Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63*(3), 506-516. https://doi.org/10.1207/s15327752jpa6303_8

Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality, 69*(6), 847-879. https://doi.org/10.1111/1467-6494.696167

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychol Assess, 12*(1), 102-111. https://doi.org/10.1037//1040-3590.12.1.102

Soto, C. J., & John, O. P. (2017a). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117-143. https://doi.org/10.1037/pspp0000096

Soto, C. J., & John, O. P. (2017b). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69-81. https://doi.org/10.1016/j.jrp.2017.02.004

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*(4), 718-737. https://doi.org/10.1037/0022-3514.94.4.718

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin, 134*(1), 138-161. https://doi.org/10.1037/0033-2909.134.1.138

Vedel, A. (2016). Big Five personality group differences across academic majors: A systematic review. *Personality and Individual Differences, 92*, 1-10. https://doi.org/10.1016/j.paid.2015.12.011

Vedel, A., Gøtzsche-Astrup, O., & Holm, P. (2018). The Danish IPIP-NEO-120: A free, validated five-factor measure of personality. *Nordic Psychology*, 1-16. https://doi.org/10.1080/19012276.2018.1470553

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect : The PANAS Scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070. https://doi.org/10.1037/0022-3514.54.6.1063

Table 1. Reliability coefficients and intercorrelations of the Danish BFI-2 domains

	Reliability		Intercorrelations			
Domain	Alpha	Retest	E	A	C	N
Extraversion (E)	.88/.83	.87*				
Agreeableness (A)	.85/.81	.85	**.27/.24**			
Conscientiousness (C)	.92/.85	.92	.15/**.15**	.14/**.16**		
Negative Emotionality (N)	.92/.89	.89	**-.35/-.22**	**-.41/-.21**	-.01/-.11	
Open-Mindedness	.88/.82	.91	**.29/.13**	**.24/.16**	.04/-.05	.07/.06
Mean	***.89/.84***	***.89***				

Note. Values left of each forward slash are for the Study 1 T2 sample ($N = 101$); values right of each forward slash are for the Study 2 sample ($N = 287$). Mean reliabilities are bolded and italicized. Correlations significant at $p < .05$ are bolded. Retest = Eight-week retest reliability in the T2 sample based on items with greatest possible overlap across T1 and T2 in Study 1. * = Retest performed excluding item 26, as the first version of this item tested at T1 in Study 1 was unrelated to other items. See Table 2 in Soto and John (2017a) for corresponding values of the original measure.

Table 2. Reliability coefficients and intercorrelations of the Danish BFI-2 facets

	Reliability		Extraversion			Agreeableness			Conscientiousness			Negative Emotionality			Open-Mindedness	
Model	Alpha	Retest	Soc.	Ass.	Ene.	Com.	Respe.	Tru.	Org.	Pro.	Respo.	Anx.	Dep.	Emo.	Int.	Aes.
Sociability	.86/.82	.91														
Assertiveness	.81/.78	.79	***.51/.43***													
Energy Level	.78/.75	.78*	***.51/.45***	***.48/.21***												
Compassion	.75/.75	.84	.16/**.20**	.09/.02	**.32/.31**											
Respectfulness	.68/.62	.66	.17/**.13**	.04/**-.12**	**.24/.27**	***.67/.53***										
Trust	.79/.63	.76	**.30/.20**	.02/.01	**.31/.30**	***.50/.45***	***.39/.52***									
Organization	.92/.87	.92	-.05/-.07	.04/.10	.12/-.01	.16/.07	.10/.05	**-.23**/-.11								
Productiveness	.85/.73	.86	.09/**.12**	**.24/.30**	**.39/.29**	**.30/.23**	**.20/.20**	-.02/.05	***.66/.47***							
Responsibility	.73/.60	.80	-.05/-.02	.07/**.16**	**.28**/.04	**.40/.30**	**.42/.28**	.00/.07	***.67/.45***	***.63/.53***						
Anxiety	.81/.76	.85	**-.32**/-.08	-.17/**-.18**	**-.26/-.12**	-.18/.10	**-.37/-.15**	**-.45/-.25**	.13/.07	.00/-.04	.14/.04					
Depression	.83/.80	.82	**-.38/-.19**	**-.38/-.26**	**-.50/-.31**	-.16/-.11	**-.31/-.27**	**-.38/-.29**	-.01/-.10	**-.21/-.27**	-.08/**-.17**	***.72/.62***				
Emotional Volatility	.84/.81	.84	-.15/-.01	-.06/**-.13**	-.07/-.03	-.14/.10	**-.42/-.20**	**-.26/-.18**	.00/.01	-.05/**-.15**	-.03/**-.16**	***.71/.63***	***.65/.55***			
Intellectual Curios.	.70/.51	.79	.19/.09	**.35**/.09	**.35/.16**	**.32/.17**	.15/**.12**	**.21/.16**	-.03/**-.13**	.12/.04	-.01/-.07	-.12/.00	-.11/.08	.10/.03		
Aesthetic Sensitivity	.82/.76	.88	.05/-.03	**.20**/-.08	**.23/.12**	.13/.07	.06/.03	.14/.11	.07/-.11	.16/.01	.00/**-.15**	.05/.08	.03/**.13**	**.21**/.08	***.62/.48***	
Creative Imaginat.	.78/.75	.89	.07/.09	**.22/.16**	**.31/.20**	.15/**.18**	.14/**.12**	**.22**/.06	-.08/.01	.03/**.14**	.04/-.03	.04/-.03	-.02/-.06	.18/.04	***.54/.46***	***.60/.45***
Mean	***.80/.73***	***.83***														

Note. Values left of each forward slash are for the Study 1 T2 sample ($N = 101$); values right of each forward slash are for the Study 2 sample ($N = 287$). Mean reliabilities and within-domain correlations are bolded and italicized. Correlations significant at $p < .05$ are bolded. Retest = Eight-week retest reliability in the T2 sample based on items with greatest possible overlap across T1 and T2 in Study 1. * = Retest performed excluding item 26, as the first version of this item tested at T1 in Study 1 was unrelated to other items. See Table 3 in Soto and John (2017a) for corresponding values of the original measure.

Table 3. Correlations of the Danish BFI-2 domains with the Mini-Markers, PANAS, and PWB

			BFI-2					MM					PANAS	
Model	Mean	SD	E	A	C	N	O	E	A	C	N	O	PA	NA
BFI-2														
Extraversion (E)	3.62	.62												
Agreeableness (A)	4.00	.54	**.17**											
Conscientiousness (C)	3.58	.76	**.18**	**.22**										
Negative Emotionality (N)	2.85	.81	**-.34**	**-.30**	-.09									
Open-Mindedness (O)	3.70	.71	**.25**	**.16**	-.02	-.01								
Mini-markers														
Extraversion (E)	5.93	1.32	***.85***	**.15**	.00	**-.35**	**.19**							
Agreeableness (A)	7.57	.96	**.22**	***.75***	**.37**	**-.17**	.11	**.18**						
Conscientiousness (C)	6.41	1.51	**.23**	**.20**	***.89***	-.10	-.02	.06	**.40**					
Negative Emotionality (N)	4.11	1.27	**-.14**	**-.41**	**-.15**	***.72***	-.07	**-.18**	**-.30**	-.11				
Open-Mindedness (O)	6.64	1.08	**.28**	**.17**	**.20**	-.05	***.81***	**.16**	**.23**	**.21**	-.13			
PANAS														
Positive affect (PA)	3.26	.68	**.40**	.13	**.15**	**-.24**	**.43**	**.27**	**.20**	**.17**	**-.18**	**.46**		
Negative affect (NA)	1.86	.63	**-.22**	**-.31**	**-.24**	**.60**	-.03	**-.22**	**-.26**	**-.21**	**.54**	**-.14**	-.08	
Psychological Well-Being (PWB)	3.96	.52	**.58**	**.31**	**.41**	**-.55**	**.21**	**.46**	**.44**	**.46**	**-.36**	**.34**	**.51**	**-.38**

Note. All values are for the Study 1 T1 sample ($N = 202$). Values were calculated based on 60 items selected to be as similar as possible to the items chosen after the second wave of data collection. Convergent correlations are italicized. Absolute correlations significant at $p < .05$ are bolded. See Table 9 in Soto and John (2017a) for corresponding values of the original measure.

Table 4. Fit statistics for confirmatory factor analyses of the Danish BFI-2 items

Model	χ^2	*df*	CFI	SRMR	RMSEA (90% CI)
Extraversion					
Single domain	511.05	54	.619	.116	.172 (.158, .185)
Single domain plus acq.	510.90	53	.619	.116	.174 (.160, .187)
Pos. and neg. items	508.27	53	.621	.117	.173 (.159, .187)
Three facets	165.81	51	.904	.059	.089 (.074, .104)
Three facets plus acq.	148.75	50	.918	.060	.083 (.068, .099)
Agreeableness					
Single domain	218.07	54	.806	.070	.103 (.089, .117)
Single domain plus acq.	187.14	53	.841	.067	.094 (.080, .109)
Pos. and neg. items	190.82	53	.837	.066	.095 (.081, .110)
Three facets	152.99	51	.879	.060	.083 (.068, .099)
Three facets plus acq.	103.35	50	.937	.053	.061 (.044, .078)
Conscientiousness					
Single domain	386.95	54	.730	.097	.147 (.133, .160)
Single domain plus acq.	377.11	53	.738	.095	.146 (.132, .160)
Pos. and neg. items	385.89	53	.730	.096	.148 (.134, .162)
Three facets	162.13	51	.910	.060	.087 (.072, .102)
Three facets plus acq	134.03	50	.932	.056	.077 (.061, .092)
Negative Emotionality					
Single domain	273.51	54	.845	.061	.119 (.105, .133)
Single domain plus acq.	270.21	53	.846	.061	.119 (.106, .134)
Pos. and neg. items	267.98	53	.848	.061	.119 (.105, .133)
Three facets	79.12	51	.980	.039	.044 (.023, .062)
Three facets plus acq.	64.80	50	.990	.034	.032 (.000, .053)
Open-Mindedness					
Single domain	248.44	54	.782	.073	.112 (.098, .126)
Single domain plus acq.	248.44	53	.781	.073	.113 (.099, .128)
Pos. and neg. items	248.43	53	.781	.073	.113 (.099, .128)
Three facets	140.60	51	.899	.057	.078 (.063, .094)
Three facets plus acq.	138.18	50	.901	.057	.078 (.063, .094)

Note. CFI = Comparative Fit Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean Square Error of Approximation. See Table 8 in Soto and John (2017a) for corresponding values of the original measure.