**The Berlin Multi-Facet Personality Inventory: An IPIP-Based Measure of Big Five Personality Facets**

**Abstract**

A novel personality inventory is presented in this article, named the Berlin multi-facet personality inventory. This new instrument is an adaptation of items from the International Personality Item Pool (IPIP; Goldberg, 2006) aimed at a more concise set of personality facets. This tool has been developed to comprise a large number of non-redundant narrow facets below each of the Big Five domains. Two language versions of the same inventory have been developed (English and German) and tested for measurement invariance in order to facilitate international usability. In addition to the construction of the inventory, this work presents promising evidence for the psychometric quality of its scores in two different populations across two different studies. The inventory is freely available online.

*Keywords*: Personality, Big Five, facet models.

**The Berlin Multi-Facet Personality Inventory: An IPIP-Based Measure of Big Five Personality Facets**

Over the last decades, the Five Factor Model (Costa et al., 1992) as well as the Big Five (Goldberg, 1990) have become widely accepted models for describing general attributes of personality[[1]](#footnote-1). In both cases, personality is conceived as a hierarchical model which describes individual differences in personality at the dispositional level: One of the most basic, universal, biologically-influenced, and stable layer of inter-individual differences in behavior, cognition, and feeling (McAdams et al., 2006). This hierarchical conception is relevant to acknowledge behavior from the most specific (nuances) to the more general (domains) differences in personality, through a varying number of mid-level characteristics (facets). Most of the research concerning criterion validity of scores from Big Five inventories has focused on the covariation between Big Five scores and relevant external outcomes. However, specific dispositional characteristics captured on the facet level might be of utility to provide more complex descriptions of individuality and to predict life outcomes to a major extent (Lounsbury et al., 2002; Paunonen et al., 2001; Ziegler et al., 2014; Ziegler et al., 2010; Kretschmar et al., 2018). Unfortunately, the number and nature of facets below the Big Five domains is far from being consensual. In fact, many different sets of facets have been proposed. One potential reason for this proliferation could be that many facet-level models were developed as an elaboration or extension to an existing domain level measure. This ad-hoc inception has the disadvantage of potentially limiting the search space for possible facets. The current research project was conducted in order to overcome such limitations and develop a facet set spanning an extensive behavior space.

**Different** **Facet Models**

As outlined above, there are a number of models that include a facet structure below the five broad domains. Among them, probably the most widely known is the one proposed by Costa and McCrae (1995), the NEO-PI-R model, which defines six facets per domain. Other popular models include the Big Five Inventory 2 (BFI-2; Soto et al., 2017), the IPIP (Goldberg et al., 2006), and the HEXACO model (Lee et al., 2016; which assumes six broad domains). A table is available in the annex (*Table A.1.*), providing an overview of these different elaborations, and listing psychometric information such as internal consistency estimates and correlations with external constructs. Other models have also been developed, such as the Faceted Inventory of the Five-Factor Model by Watson et al. (2017), or the Big Five Aspect Scales by DeYoung et al. (2007).

The abundance of proposed facets related to each of the Big Five domains evidences the reality of a topic that is not consensual and that gives rise to many different elaborations. Despite these differences, however, there is still some degree of overlap.

Soto et al. (2009) inspected the convergence between the NEO-PI-R and the first version of the BFI, suggesting that two constructs per domain were measured at the facet level by both inventories. The constructs defined by Soto et al. (2009) were: *Altruism* and *Compliance* for agreeableness; *Anxiety* and *Depression* for neuroticism; *Order* and *Self-Discipline* for conscientiousness; *Assertiveness* and *Activity* for extraversion; and *Aesthetics* and *Ideas* for openness. Likewise, the existence of such “core” constructs was also suggested by DeYoung et al. (2007), in what they termed aspects. Even though both contributions’ labels vary, they have a substantial degree of similarity in terms of content. Furthermore, these core constructs are present not only in the models which Soto et al. (2009) analyzed, but also in all models listed in *Table A.1*. Some of these constructs are explicitly covered (e.g. *Anxiety*), while others are implicitly encompassed, especially when given attention to the items’ content (e.g. *Liveliness* in HEXACO resembles the “core” construct *Activity*, present in all other instruments).

While the majority of models based on the Big Five include these “core” constructs, there is still an abundance of other constructs which could be termed “peripheral”, whose inclusion is more variable. Soto et al. (2017, page 118) referred to this diversity of constructs by saying that the Big Five dimensions *“can be conceptualized and assessed more broadly or more narrowly”*, either by focusing on central or on peripheral facets, depending on the interest of research. It is due to this multiplicity of peripheral constructs that an effort to convey different models would be beneficial, in order to obtain a comprehensive inventory which subsumes a maximum possible of these peripheral facets, as well as including the essential core facets.

An important step towards building such inventory is to ensure that the proposed set of facets predict consequential outcomes. One of the most classical approaches to relate the measured constructs with external constructs is to define a nomological network between the personality traits and external outcomes (Cronbach et al., 1955). Nomological networks can be drawn from a hierarchical perspective too. Relying on dimensions to describe behavior and to predict external outcomes can benefit from ease of interpretability. However, the domain level is sometimes too distal to depict behavioral mechanisms underlying personality-to-outcome associations thoroughly. Conversely, using nuances to predict behavior might yield a stronger predictive power (Seeboth et al., 2018), as specificity to situations and contexts is enhanced (Ziegler et al., 2016). Nonetheless, using nuances in the prediction of external outcomes can have the disadvantage of dealing with extreme complexity. Facets are by definition in a middle ground between nuances and dimensions, in a compromise between specificity and sensitivity in the bandwidth-fidelity dilemma. This aggregation could satisfy the specificity of predictions while at the same time enhance the ease of interpretability of personality-to-outcome associations.

## Facets associated with consequential outcomes

The question of whether narrow measures have a superior predictive power over broad measures has enjoyed a lively debate. Research summarized below suggests an advantage for scores derived from more narrow measures. Nonetheless, other lines of research point at the opposite direction (Salgado, 2017; Chen, 2012). It falls out of the scope of this study to provide evidence which could fuel this debate further. Still, we believe that the inventory presented here will be useful to forthcoming contributions on this topic.

The following section provides an overview of evidence for relations between domains, facets, and three consequential outcomes that are of interest to researchers.

**Satisfaction with Life**

One of the outcomes that has been largely evidenced to be predicted by personality is satisfaction with life (SWL). There is a consensus on signaling emotional stability and extraversion as the most important domains when predicting SWL (Heller et al. 2004; and Steel et al. 2008). At the facet level, Steel et al. (2008) and Schimmack et al. (2002) identified the NEO facets of *Depression* and *Positive emotions* as the most important predictors, with medium sized regression coefficients. In line with these findings, we hypothesize that the set of facets which measure emotional stability and extraversion in our inventory should significantly correlate with SWL, with a moderate to big effect size.

**Academic Performance**

Another relevant outcome that has being typically predicted by personality is academic achievement. Conscientiousness has been recognized as the strongest Big Five domain to predict this outcome (Poropat, 2009; 2014). At the facet level, De Fruyt et al. (1996) hypothesized that facets of conscientiousness related to volition would be more strongly related with academic achievement. There is a collection of research that is consistent with this idea, linking academic performance with facets such as *Achievement-striving* (O’Connor et al., 2007; Gray et al., 2002) or *Work drive* (Lounsbury et al., 2002). Nonetheless, other conscientiousness facets more related to duty or moral behavior have been found to predict GPA: for instance *Self-discipline* (O’Connor et al., 2007; Gray et al., 2002), or *Dutifulness* (O’Connor et al., 2007). The relation of academic achievement with openness at the domain level has been more variant, in part because the facets of openness can be related in opposite directions with this outcome. Paunonen et al. (2001) found that the openness facet *Understanding* correlates with academic achievement. Noftle et al. (2007) identified a set of NEO-PI-R and HEXACO openness facets which predicted academic achievement (the HEXACO facets of *Aesthetic*, *Inquisitiveness*, *Creativity* and *Unconventionality*, plus the NEO-PI-R facets of *Fantasy*, *Aesthetics*, *Feelings* and *Ideas*, also see Ziegler et al., 2010). In line with these findings, we hypothesize that the facets which entail the conscientiousness domain in our inventory would consistently predict academic performance, with a small to moderate effect size, and that openness to experience will yield a mixed pattern at the facet level.

**Academic Absenteeism**

Personality has also been reported as a robust predictor of absenteeism, with especial attention to inverse correlations between this outcome and the domain conscientiousness (Judge et al., 1997; Ones et al., 2003). Some specific facets of conscientiousness have been highlighted, like *work drive* (Lounsbury et al., 2004) or *need for achievement* (﻿Wegge et al., 1993). Chamorro-Premuzic et al. (2003) as well as ﻿Furnham et al., (1995) found significant direct correlations with openness, while direct correlations with extraversion were reported by ﻿Judge et al. (1997). The relation of abseentism with the personality at the domain level has been vaguer. We expect that this ambiguity could be resolved by switching the focus to the facet level.

## This research

Attending to the multiplicity of facet models available in the personality literature and the resurgence of narrow constructs as relevant units of analysis to describe personality and to predict important life outcomes, we strove to develop a comprehensive measure of personality facets which covers the content of the Big Five domains to a major extent.

This novel measure is an adaptation of items from the International Personality Item Pool (IPIP; Goldberg, 2006) to a more concise measure of personality traits with a focus on narrow facets. An antecedent to this study can be found in MacCann et al. (2009), where part of the stimuli set presented here was used in order to design a measure of conscientiousness containing a maximum number of relevant facets. This research extends MacCann et al. (2009) to all Big Five domains. Furthermore, the inventory presented here has been developed with the scope of being open source, by making it available to researchers and practitioners at no cost; and with the intention of facilitating internationally usage by testing its applicability in two different cultures. Furthermore, the same taxonomy has been used in a different study, albeit with a different set of items (Ziegler et al., 2019). In that study, the authors compare this taxonomy with other important taxonomies of Big Five personality.

It has to be noted that proposing such a higher number of facets requires evidence to support the usefulness of each facet (Siegling et al., 2015). At the same time, it is also clear that such an endeavor must be considered as permanent work in progress. Within this paper we will test the idea that facets outperform domain scores with regard to amount of criterion variance explained. Moreover, we will use the approach suggested by Rosenbusch et al. 2020, and test the content similarity of our facet items with scales uploaded in a data bank. Importantly, this data bank contains the IPIP scales and keys. This way we can avoid jingle-jangle fallacies.

In order to evidence its applicability among cultures, this research has been conducted in two studies using two independent samples from two different countries: the USA and Germany. In the first study, using the US sample, we empirically defined a facet model, selected items, and tested the facets’ psychometric properties with regard to factorial validity evidence, internal consistency, and test-criterion correlations. In Study 2, we replicated the findings with a German sample, and further tested the measurement equivalence of the suggested models. Methods and results sections are presented separately for each study. General discussion and conclusion sections are provided at the end of this manuscript.

## Study 1

### Methods

#### Participants. This sample consisted on 722 American undergraduate students (59.30% male) who gave voluntary acceptance to their inclusion in this research. Their mean age was 21.60 years (SD = 5.90), and 59% of them were female. Students were e-mailed a link to a computerized assessment battery that included the items as well as several other tests not reported in this paper. The data set was randomly split into two equally sized subsamples. Both subsamples were matched in relation to missing values, outliers, and extreme values. Subsample 1’s mean age was 21.80 years (SD= 6.30), subsample 2’s mean age was 21.50 years (SD=5.60).

#### Measures. Altogether, 525 items from the IPIP were used in this study, as indicators of the Big Five domains. The IPIP is an open-source database of personality items, launched in 1996 and containing over 2000 items (Goldberg et al., 2006). Participants were asked to provide self-ratings of personality items on a 5-point rating scale, ranging from 1 (“Not all like me”) to 5 (“Very much like me”).

The item set used originates from a complex item reduction conducted before. Here, all IPIP items were rated regarding their prototypicality for a Big Five domain. Based on these ratings, items for the current study were selected. More details can be found in MacCann et al. (2009).

##### Satisfaction with life (SWL). Measured with a 5-item composite defined in Diener et al. (1985), in a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Items included are: a) “In most ways my life is close to ideal”, b) “The conditions of my life are excellent”, c) “I am satisfied with my life”, d) “So far I have gotten the important things in my life”, and e) “If I could live my life over, I would change almost nothing”. Psychometric properties have been reported excellent. In our sample, the reliability estimate was α = 0.88.

##### Graded point average (GPA). To measure academic performance, participants reported their GPA scores at the end of high school.

##### Absences: Participants reported an estimation of days that were absent from college without justification. This was an item extracted from a larger set of student social behaviors indicators (MacCann et al., 2009). Absences were log transformed prior to analyses, as proposed by Lounsbury et al. (2004).

#### Statistical Analyses

Analyses in this study were conducted applying a split-sample method to separate its exploratory and confirmatory phases. An Exploratory Factor Analysis (EFA) was conducted with subsample one, in order to identify the number of components underlying the personality items in our dataset. Subsample two was used in order to calculate reliability by means of internal consistency, to conduct tests of unidimensionality with Confirmatory Factor Analysis (CFA), and to obtain evidence of structural validity with Exploratory Structural Equation Modelling (ESEM).

Finally, a nomological network was built with subsample two, in order to examine preliminary evidence of construct validity of our proposed facet model. We elaborated the following hypotheses regarding our nomological network.

* H1. SWL will be predicted by facets of emotional stability mimicking NEO-PI-R *depression*, and facets of extraversion covering *positive emotions*, with a big to moderate effect size, in line with Schimmack et al. (2002). Emotional stability and extraversion will be most important domains in the personality-SWL association.
* H2. Conscientiousness will be associated with academic achievement with a small to moderate effect size. Openness will entail facets with positive effects and facets with negative effects on GPA scores.
* H3. Conscientiousness will yield the strongest associations with abseentism at the domain level, and facets tapping volitional components such as *goal orientation* or *wish to work* will outstand. Some specific facets of openness and of extraversion will also be significantly associated with abseentism. Overall, the facet level will provide a clearer picture to predict academic abseentism from personality than the domain level.

A more detailed description of the statistical analysis can be found in the annex.

## Results

##### EFA. Exploratory analysis revealed that the domains could be structured into eight to eleven facets. Model fit information for the EFA procedure, as well as Eigenvalues, and results from the MAP and PA tests, are presented in the annex. To ensure the homogeneity of the facets and to reduce the risk of cross domain loadings, items with factor loadings of less than .30 and with non-central content to the domain in question were eliminated (Ziegler, 2014).

Eight facets were retained for the domain agreeableness, after two were eliminated due to weak loadings and clusters whose content was elusive. These facets were named *Appreciation*, *Integrity*, *Low competitiveness*, *Readiness to give feedback*, *Search for support*, *Compliance*, *Genuineness,* and *Altruism*. Items corresponding to these and the following facets can be found in the annex.

Conscientiousness consisted of nine facets after one facet with factor loadings below .30 was excluded, these were: *Dominance*, *Persistence*, *Self-discipline*, *Task planning*, *Goal orientation*, *Carefulness*, *Orderliness*, *Wish to work* (to capacity), and *Productivity*.

Extraversion was formed by nine facets. A new facet (*Energy*) was added in order to tap the physical component of extraversion, which was missing in the eight-facet solution the EFA suggested. These facets were labelled *Sociability*, *Readiness to take risks*, *Wish for affiliation*, *Positive attitude*, *Forcefulness*, *Communicativeness*, *Humor*, *Conviviality,* and *Energy*.

Neuroticism (interpreted here as emotional stability) consisted of seven facets. One facet was dropped due to poor interpretability and was therefore not included in the subsequent analyses. The final set of facets were named *Equanimity*, *Confidence*, *Carefreeness*, *Mental balance*, *Drive*, *Emotional robustness,* and *Self-attention*.

Openness to experience comprised nine facets. One facet was identified as a method factor and eliminated, as it solely contained negatively formulated items and no coherent underlying trait could be identified. Furthermore, an extra facet was added (*Intellect*), as it was not present in the EFA solution and represents a core construct in other important facet models. The final set of facets of the openness domain were named *Creativity*, *Wish for variety*, *Open-mindedness*, *Interest in reading*, *Aesthetics*, *Wish to analyze*, *Willingness to learn*, *Sensitivity,* and *Intellect.*

### *Reliability.* Reliability estimates for each of the facets and all domains were obtained using Cronbach’s (Cronbach, 1951) and McDonald’s (McDonald, 1991). 95% C.I. estimates of McDonald’s for the domains were: agreeableness ranged from 0.85 to 0.91, conscientiousness ranged from 0.83 to 0.88, opennessranged from 0.91 to 0.94, emotional stability ranged from 0.90 to 0.93, extraversion ranged from 0.89 to 0.92. All in all, reliability coefficients were at least good ( > 0.8) for all the domains, and at least acceptable ( > 0.7) for the majority of the facets (60%). Only one facet had poor internal consistency (*Altruism*, . All internal consistency estimates for the facets can be found on *Table 1*.

### *CFA*. Confirmatory factor analysis was applied to each of the facets identified in the previous step, using the second American-based sample. All measurement models fitted well according to goodness-of-fit indices. The fit information of three facets was not available as these models were reflected by only three indicators and therefore saturated (*energy*, *self-attention*, and *intellect*). Goodness of fit estimations for each facet are available in *Table 1.*

### *ESEM*. The final ESEM model was constructed after removing four facets that did not significantly load in their intended domain: *sensitivity* (expected to load on openness), *search for support* and *readiness to give feedback* (expected to load on agreeableness), and *readiness to take risks* (expected to load on extraversion). Furthermore, two residuals were allowed to be correlated after inspection of modification indices: Emotional stability’s facet *drive* was correlated with conscientiousness’ facet *persistence*, as well as extraversion’s facet *forcefulness* with conscientiousness’ facet *dominance*. The addition of these correlated residuals was consistent with the facet’s content and revealed that a significant amount of specific variance was still present in the facets. The resulting model comprised 38 facets, all of them with significant loadings in their intended domains. It yielded a model fit of (df) = 1614.8 (521), CFI = .83, RMSEA = .074, SRMR = .043. We consider this model fit as sufficient to approximate our data according to both RMSEA and SRMS, which indicate that the size of the residuals was not substantial. The CFI value was lower than the usual cut-offs, probably due to facet specific variance which was not accounted for in the model. The standardized loadings of the facets in their intended domain can be found in *Table 1*. As it is usual in ESEM procedures, some facets presented significant cross-loadings in other domains. These significant cross loadings ranged from λ= 0.12 to λ= 0.68.

------------------ Insert table 1 around here ------------------

### *Nomological network.* Assumptions from a nomological network with our proposed set of facets and three external criteria were tested in order to provide evidence of criterion validity. *Table A.3* in the annex summarizes these findings and highlights that, overall, the hypotheses outlined in the methods section were met. H1 stated that emotional stability and extraversion should be the domains which had bigger associations with SWL. Indeed, both domains showed the highest correlations and their models accounted for the most variance explained by the predictors. At the facet level, *mental balance* and *positive attitude* (both β > 0.5) were strongly associated with the criterion, also in line with what was hypothesized in H1. H2 stated that conscientiousness would be associated with GPA with a medium to small effect size and that openness would yield a mixed pattern of association at the facet level. Conscientiousness was, in line with previous research, the domain with higher associations on SWL, with a correlation of *r* = 0.26. Three openness facets were significantly associated with the outcome: *creativity* was inversely associated (β = -0.14), while *interest in reading* and *intellect* correlated with it directly (β = 0.12 and 0.17). Thus, our nomological network showed the expected properties regarding H2. H3 stated that conscientiousness should be the domain yielding highest associations with abseentism, and that conscientiousness’ facets related to volitional aspects would highlight this association. Indeed, conscientiousness had the strongest associations with abseentism (in an inverse relation, *r* = -0.28), and two facets related to volition, *task-planning* and *productivity*, were significantly associated with this outcome. Furthermore, H3 stated that the relation of abseentism with other domains will be clearer at the facet level. In this line, we have only found modest associations at the domain level, but some facets like *genuineness*, *energy*, *willingness to learn*, *humor* or *drive* were significantly associated with the outcome, thereby confirming H3.

*Content-analysis.* The linguistic similarity check (available in the supplemental materials) revealed that some of the facets we propose are already suggested in the IPIP keys. At the same time, there are also several facets that are not in those keys. In order to avoid jingle jangle fallacies, we included the results of our similarity check in the online supplement.

## Study 2

### Procedure and Participants

Study two was conducted with an independent sample of 387 German speakers (49.10% male) with a mean age of 45.60 years (SD = 17.50), representative for the German working population with regard to age, gender, and education level. The data was collected in a test center.

### Measures

For the German version of the presented tool, the IPIP items selected in Study 1 were translated and back-translated by bilingual experts. Non-matching back-translations were flagged as inadequate and were further adapted by the same experts.

### Data analysis

Similarly as proceeded in Study 1, internal consistency and structural validity by means of one CFA model per facet and one ESEM model for the full inventory was conducted with the German sample. In addition, we present in this section a measurement invariance procedure used to test the equivalency of the measurement models in the two countries.

#### Measurement invariance (MI). Following tests for structural validity mirroring the procedures from Study 1, MI tests were conducted for each facet using the German and the US data. Three levels of MI were analyzed here: configural, metric, and scalar invariance. Model comparison was based on suggestions by Chen (2007). Metric invariance was accepted whenever CFI < .01, RMSEA < .015 or SRMR < .03; and scalar invariance whenever CFI < .01, RMSEA < .015 or SRMR < .01. MI for the full model was tested using ESEM.

In addition to full invariance tests, partial invariance was also tested at the facet level. Partial invariance was investigated by allowing a maximum of two factor loadings (for metric invariance) or intercepts (for scalar invariance) to differ between countries. The robust maximum likelihood estimator (MLR) was used in to fit the factor models in tests of invariance as it is simpler than categorical estimators when inspecting invariance, especially at the scalar level.

### Results

### MI at the facet level. Configural invariance was found in all facet models. This was the highest degree of invariance obtained for one facet, *Readiness to give feedback*, a facet of agreeableness. A relatively high number of facets reached partial scalar invariance (48.9%) after freeing a maximum of two intercepts in their respective models. One facet reached full scalar invariance: *Sociability*, a facet of extraversion. Furthermore, the vast majority of facets showed at least partial metric invariance between both countries (97.6%) (see Table A.5 in the annex).

As stated above, the partial invariant solutions were tested after freeing a maximum of two parameters between groups in each MI stage. As a general trend, non-invariant factor loadings of extraversion are larger in the USA sample than in the German sample, as well as positive factor loadings of openness. Conversely, negative factor loadings of openness are bigger in the German sample. Some freed factor loadings seem to deviate due to age differences between samples, as “I resist authority”, an indicator of *Dominance*, and “I am easily discouraged”, an indicator of *Persistence*, both yielding higher factor loadings in the younger sample (i.e. our German sample). At the scalar level, most non-invariant parameters of conscientiousness’ facets show higher intercepts for Germans, excepting those of *Persistence*, which are higher in the American sample. Nonetheless, *Persistence* indicators may be higher in the American sample due to age differences (“I like to take my time” and “I never give up” are examples of items belonging to this facet). In fact, many of the non-scalar indicators may be best understood by the effect of age than that of culture, especially those that differ the most; for instance: “I act impulsively when something is bothering me”, “I am easily talked into doing silly things”, “I get overwhelmed by emotions”, all of these yielding higher intercepts in the younger group.

### MI of the full model. The ESEM model with the German sample showed similar fit as with the American sample ( (df) = 1386 (521), CFI = 885, RMSEA = 0.068, SRMR = 0.035). Importantly, all facets loaded significantly in their intended domains, replicating the results of Study 1. The MI approach revealed that configural invariance was tenable in the integrated model ( (df) = 3001(1042), CFI = 0.86, RMSEA = 0.071, SRMR = 0.04). Metric invariance was not obtained, as differences in CFI were higher than 0.01 (*∆*CFI = 0.026). Partial invariance was not tested as it is not yet implemented in ESEM.

# Discussion

The personality test presented herein, named Berlin multi-facetted personality inventory (BMFPI), was developed to cover the need for a tool which maximized the coverage of facets within the Big Five framework. Starting from a large online item pool, we have developed a questionnaire which assesses 38 facets with 202 items. The selected facets cover central constructs which are present in most Big Five models that include facet levels, as well as more peripheral constructs which could help to describe individual differences in a more nuanced manner. The first evidence of reliability, construct and test-criterion validity of the set of facets has been promising, according to the results presented in this manuscript. In addition, the BMFPI has been developed to enable cross-cultural usage and to align with the principles of open accessibility, ensuring that researchers worldwide can benefit from this advance. These analyses also revealed interesting patterns of non-invariance, potentially informing cross-cultural research. For instance, higher factor loadings were found in the American sample for positively keyed items in the openness domain, whereas inversely keyed items loaded highly in German’s openness. This suggests that the indicators which reflect a low level of openness have a relatively higher influence in how this domain is defined by Germans. Also, non-invariant intercepts in conscientiousness were consistently higher in the German sample, indicating a higher baseline level in this domain for this group.

### Facet Structure

The instrument presented in this work covers all the “core” facets proposed by Soto and John (2009 & 2017). In some cases, these core constructs have been labelled similarly to the proposal of the cited authors, that is the case of *energy*, *altruism*, *compliance, order* or *self-discipline*. In some other cases, our proposed labels were different as these were defined more narrowly than in Soto and John (2009). That is the case of *mental balance* (instead of *anxiety*), *emotional robustness* (*depression*), or *artistic interest* (aesthetics). In the remainder of cases, the core constructs were represented by more than a single facet, to account for nuances in facets that we believe are more heterogeneous. This is the case of *low competitiveness* and *integrity*, which both can be thought as related to *assertiveness*; or *open-mindedness* and *wish to analyze*, which are tapping the *ideas* component of the openness domain.

Using a faceted instrument requires specific information within the facet scores (Ziegler & Bäckström). Here we provide initial evidence that the facets we propose are relevant with regard to test-criterion correlations. Moreover, based on linguistic similarity we could show that several of the facets we propose reflect facets already specified in the IPIP framework. However, the similarity check also underscores the potential of the newly proposed facets revealing low similarity with other established personality measures. Some of them may be highly specific, such as interest in reading, although recent research suggest that this facet may be central to the domain of Openness (Trapp and Ziegler, 2019). Despite these encouraging findings it is mandatory that future research further investigates the structural robustness as well as the potential for test criterion correlations of all proposed facets.

### Psychometric properties

After defining the facet structure with an independent sample, we tested the psychometric properties in terms of internal consistency and structural validity. We found good internal consistencies, with 67% of the facets ω greater than 0.70 and 95% of the facets ω > 0.60. The domains were also reliably measured, with ω ranging from 0.83 to 0.92.

Structural validity was assessed by fitting a CFA to each of the facets. Goodness of fit measures signaled that the data is consistent with the facet models, with 88% of the chi-square tests yielding non-significant results and all facets at least approximately fitting the data according to goodness-of-fit indices. These results suggest that the facets included in the BMFPI can be used independently, in case that researchers and practitioners are more interested in a specific set of facets rather than in the full Big Five picture.

The higher order structure was tested using ESEM, a method which allows to overcome the constraint of independent clusters solution usually imposed in a CFA. Although some researchers could argue that such constraints are beneficial to ensure a high degree of discriminant validity, the independent cluster solution may be too strict to model constructs that are highly inter-related, as it occurs in personality traits. A control measure to deduce a reasonable degree of convergent and discriminant validity in the ESEM solution is to verify that the primary factor loadings are distinct from zero (i.e. that they are statistically significant), and that the proportion of significant cross-loadings is not excessive. All of this was the case for the current data in both languages. Cross-loadings were as expected and, as we used facet scores instead of sum-scores, these were incorporated into further analyses such as the nomological network.

### Association with external constructs

We have tested different hypotheses which aimed to replicate previous findings on the interplay between personality and SWL, academic performance, and school absenteeism. Overall, the scores derived here have shown similar test-criterion-correlations as scores from other personality instruments, at least regarding the external outcomes that were included in our measures. Confirming our first hypothesis we found that *positive attitude* (which is similar to *cheerfulness*) and *mental balance* (akin the inverse of *depression*) were associated with the outcome with a medium effect size (β = .51 and β = .59, respectively). Emotional stability and extraversion were, as expected, the domains which showed the strongest links to SWL. The other domains did not account for a big piece of variance of SWL ( ranging from 0.08 and 0.13), although we did find some interesting associations at the facet level. For instance, *integrity*, *good faith*, *persistence*, *confidence*, *open-mindedness* and *intellect* were significantly linked with SWL, highlighting the usefulness of a rich set of facets when inspecting associations with external outcomes.

Our second hypothesis stated that conscientiousness would be associated with academic achievement with a small to moderate effect size, and that openness would result in a heterogeneous structure of direct and inverse effects at the facet level. Our results were in line with the hypothesis. Conscientiousness’ sum score yielded a correlation of r = .23 with high school GPA, while at the facet level *goal orientation* (β = .15) was the only construct predicting the outcome. Openness was also related to GPA, and as hypothesized some facets were positively related to academic achievement *(Interest in Reading*, β = .12; *Intellect*, β = .17), while some others were related negatively (*Creativity*, β = -.14). Furthermore, two facets of agreeableness were linked with high school GPA, they were *low competitiveness* (β = -.13) and *genuineness* (β = .14). The last facet that was significantly associated with the outcome was emotional stability’s *equanimity* (β = .13).

The third hypothesis stated domains would not be directly linked with absenteeism in high school, but that some facets would picture this association more clearly. In line with this hypothesis, all were modest ( ranging from 0.03 to 0.1), but some specific facets were significantly associated with the outcome. These were agreeableness’ *genuineness* (β = -.15), conscientiousness’ *task planning* (β = -.14) and *productivity* (β = -.13), extraversion’s *energy* (β = -.18) and *humor* (β = .13), emotional stability’s *drive* (β = -.14), and openness’ *willingness to learn* (β = -.15). These results highlight that, in order to better predict educational absenteeism, researchers should focus on narrow constructs instead of on broader domains.

All in all, our nomological network suggest a reasonable degree of construct validity for the BMFPI. Furthermore, our results suggest that switching the focus from a domain perspective towards a facet perspective may increase the strength of associations between personality scores and external criteria.

### International usage

This instrument aims to be usable in different countries to promote internationalization of individual differences research. To this aim, it has been tested in two different languages, with samples gathered from two countries in two different continents. We have applied measurement invariance techniques to test the extent to which both versions of the inventory are equivalent. At the facet level, all of the facets yielded configural invariance, and the vast majority of the facets reached partial metric invariance. Scalar invariance was attained in one facet. Nonetheless, it is important to notice here that the two populations were very different in terms of age (effect size of the difference *d* = 1.83). Given that personality traits’ structure and mean levels change within the lifespan of individuals, this age difference might have affected our invariance tests. For instance, Roberts et al. (2006) reported in their meta-analysis that the mean levels of *Dominance* increase from late childhood into late adolescence, reaching a plateau at an approximate age of 35. Our results are in line with this finding. The facet *Dominance* is not scalar invariant, and only partial metric invariance could be established. Future studies should investigate normative change as a possible confounder in the results presented in this paper, as our two samples’ mean age differ significantly. Despite this mean age difference, the degree of measurement invariance obtained by both versions of the inventory has been satisfying, indicating that cross-cultural research is feasible with the inventory presented here.

### Limitations

Given that this study is the first outlet using the BMFPI, its psychometric properties are still preliminary and subject to replication. More studies are needed with heterogeneous samples to test whether the found structure can be retrieved from other populations. We should bear in mind that some facets which were a priori defined had to be excluded from the final proposal due to deficits in their integration within the Big Five framework, this problem could arise in other populations.

One important limitation was sample size with regard to ESEM models which could not be fitted using a second order structure. Instead, we were forced to use the factor scores from previous facet models as indicators of the domains, as the model did not converge otherwise. By doing so, we had some specific facet variance left outside the ESEM model which resulted in lowered values for the CFI.

Currently, the BMFPI is only available in two languages. We hope to see a bigger dissemination through different cultures in order to extent its international usability. Although personality traits seem to be universal, a certain degree of cultural variation does exist. Further adaptations of this inventory will reveal the extent to which the chosen indicators relate to the proposed set of facets in countries distinct from the USA or Germany.

Furthermore, the demographics of the two samples differ considerably. Nonetheless, the promising results on the measurement invariance of the scales across the two countries reveal that the instrument measures the constructs similarly in both of them.

Finally, several psychometric properties of this instrument have not yet been tested, such as temporal stability or congruent and discriminant validity with other Big Five instruments. However, a previous study has examined these with the same taxonomy, although using a different set of items (Ziegler et al., 2019). Future research should also examine a more extensive nomological network, especially attending to the domains Agreeableness and Openness.

# References

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, 80, 219–251.

Costa, P. T., & McCrae, R. R. (1992) Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653-665

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, *64*(1), 21–50.

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

Cronbach, L., J., Meehl, P., E., (1995). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

De Fruyt, F., & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality*, *10*(5), 405–425.

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.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality*, *49*(1), 71–75.

Furnham, A., & Medhurst, S. (1995). Personality correlates of academic seminar behaviour: A study of four instruments. *Personality and individual differences*, 19(2), 197-208.

Gray, E., & Watson, D. (2002). General and Specific Traits of Personality and Their Relation to Sleep and Academic Performance. *Journal of Personality*, *70*(2), 177–206.

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. Journal of personality and social psychology, 59(6), 1216.

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*(1), 84–96.

Heller, D., Watson, D., & Ilies, R. (2004). The role of person versus situation in life satisfaction: A critical examination. *Psychological bulletin*, *130*(4), 574.

Judge, T. A., Martocchio, J. J., & Thoresen, C. J. (1997). Five-Factor Model of Personality and Employee Absense. *Journal of Applied Psychology*, *82*(5), 11.

Lee, K., & Ashton, M. C. (2016). Psychometric Properties of the HEXACO-100. *Assessment*, *1-15*.

Lounsbury, J. W., Sundstrom, E., Loveland, J. L., & Gibson, L. W. (2002). Broad versus narrow personality traits in predicting academic performance of adolescents. *Learning and Individual Differences*, *14*(1), 67–77.

Lounsbury, J. W., Steel, R. P., Loveland, J. M., & Gibson, L. W. (2004). An investigation of personality traits in relation to adolescent school absenteeism. *Journal of Youth and Adolescence*, *33*(5), 457–466.

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of Conscientiousness. *Learning and Individual Differences*, *19*(4), 451–458.

McAdams, D. P., & Pals, J. L. (2006). A new Big Five: Fundamental principles for an integrative science of personality. *American Psychologist*, *61*(3), 204–217.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552–566

McDonald, R. P. (1999). *Test theory: A unified treatment*. psychology press.

Noftle, E. E., & Robins, R. W. (2007). Personality Predictors of Academic Outcomes: Big Five Correlates of GPA and SAT Scores. *Journal of Personality and Social Psychology*, *93*(1), 116–130.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: a meta-analysis of integrity tests. *European Journal of Personality*, *17*(S1), S19–S38.

O’Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971–990.

Paunonen, S. V., & Ashton, M. C. (2001). Big Five Predictors of Academic Achievement. *Journal of Research in Personality*, *35*(1), 78–90.

Poropat, A. E. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin*, *135*(2), 322–338.

Poropat, A. E. (2014). A meta-analysis of adult-rated child personality and academic performance in primary education. *British Journal of Educational Psychology*, *84*(2), 239–252.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological bulletin*, 132(1), 1.

Rosenbusch, H., Wanders, F., & Pit, I. L. (2020). The Semantic Scale Network: An online tool to detect semantic overlap of psychological scales and prevent scale redundancies. *Psychological methods*, *25*(3), 380.

Salgado, J. F. (2017). Bandwidth-fidelity dilemma. *Encyclopedia of Personality and Individual Differences*, eds Zeigler-Hill V., Shackelford TK, . Springer International Publishing, 1-4.

Schimmack, U., Diener, E., & Oishi, S. (2002). Life-satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. *Journal of Personality*, *70*(3), 345–384.

Seeboth, A., & Mõttus, R. (2018). Successful explanations start with accurate descriptions: Questionnaire items as personality markers for more accurate prediction and mapping of life outcomes. *Journal of Personality, 32(3), 186-201*.

Siegling, A. B., Petrides, K. V., & Martskvishvili, K. (2015). An Examination of a New Psychometric Method for Optimizing Multi‐Faceted Assessment Instruments in the Context of Trait Emotional Intelligence. European Journal of Personality, 29, 42-54.

Soto, C. J., & John, O. P. (2009). 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*(1), 84–90.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, 113(1), 117.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological bulletin*, *134*(1), 138.

Trapp, S., & Ziegler, M. (2019). How openness enriches the environment: Read more. *Frontiers in psychology*, *10*, 1123.

Watson, D., Nus, E., & Wu, K. (2017). Development and Validation of the Faceted Inventory of the Five-Factor Model. *Assessment,* 26 (1), 17-44

Wegge, J., & Kleinbeck, U. (1993). Motivationale Faktoren betrieblicher Fehlzeiten: zum Einfluß leistungs-und anschlußthematischer Variablen auf die Abwesenheit am Arbeitsplatz. *Zeitschrift für experimentelle und angewandte Psychologie*, 40(3), 451-486.

Ziegler, M. (2014). Comments on item selection procedures. *European Journal of Psychological Assessment*, 30, 1-2.

Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting Academic Success with the Big 5 Rated from Different Points of View: Self-Rated, Other Rated and Faked. *European Journal of Personality*, *24*(July), 341–355.

Ziegler, M., & Brunner, M. (2016). Test standards and psychometric modeling. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), The Springer series on human exceptionality. Psychosocial skills and school systems in the 21st century: Theory, research, and practice (p. 29–55). *Springer International Publishing*.

Ziegler, M., Horstmann, K. T., & Ziegler, J. (2019). Personality in Situations: Going beyond the OCEAN and Introducing the Situation Five. Psychological Assessment. Advanced online publication. http://dx.doi.org/10.1037/pas0000654

1. Often the terms are even used synonymously, which is why we will refer to the Big Five from here on. [↑](#footnote-ref-1)