Edinburgh Research Explorer

Towards a taxonomy of personality facets

Towards a taxonomy of personality facets

Paul Irwing[1*], David J. Hughes[1*], Alexander Tokarev[1], Tom Booth[2]

[1]Alliance Manchester Business School, The University of Manchester.

[2]Department of Psychology, The University of Edinburgh.

*The first two authors contributed equally to this research

The Facet-MAP Version 1 – self-report form, including items, and instructions for use is here:

https://osf.io/t2ujq/?view_only=2b2900fed17a4ed1856c25bf74472678

Correspondence concerning this article should be addressed to: Prof. Paul Irwing, Alliance Manchester Business School, The University of Manchester, Room 7.005, Booth Street West, Manchester M15 6PB, Tel: +44 (0) 161 832-3872, Email: paul.irwing@manchester.ac.uk

Towards a taxonomy of personality facets

Abstract

We describe three studies that together provide a first approximation to a comprehensive taxonomy of unique personality facets. In Study 1, we semantically sorted, removed synonyms, and factor analysed 1,772 personality items taken from seven major omnibus personality inventories and four narrow inventories. Study 1 identified 61 base facets. In Study 2, we conducted a systematic review of the literature to identify facets missing from the 61 base facets. We identified 16 novel facets. We then created standardised, open access items for the 77 facets. In Study 3, we administered the items to a novel sample (N = 1,096) and assessed the psychometric properties of the facets. The ultimate result was 70 personality facet scales that are open-access, psychometrically robust, unidimensional, and discriminant. We call this inventory the Facet-level Multidimensional Assessment of Personality or Facet MAP, version 1. The Facet MAP contains scales equivalent to almost all scales present in major personality inventories, and in most cases, many more as well. As the Facet MAP develops, we hope it will eventually provide a comprehensive taxonomy of personality facets, which will prove useful in reducing construct proliferation and facilitating numerous avenues of important personality research.

Keywords: Personality Traits; Factor Analysis; Personality Structure; Personality Scale; Five Factor Model.

Towards a taxonomy of personality facets

Taxonomies of human personality traits describe phenotypic patterns of thinking, feeling, and behaving. Most personality trait taxonomies and inventories are hierarchically structured, building from individual items (nuances), through narrow traits (facets), to broader traits (dimensions/domains/factors). Figure 1 contains an illustrative example of a hierarchical model of personality based on the Big Five/Five-Factor Model tradition. Most taxonomic research has focused on broad trait domains with comparatively little having examined the taxonomic structure of narrower traits (Schwaba, et al., 2020; Soto & John, 2017; Ziegler & Backstrom, 2016). Where taxonomic research has examined narrow traits, it has mostly been piecemeal in nature, with "little systematic research yet to delineate an empirically based and comprehensive model of personality facets" (Mõttus et al. 2020, p. 1177). In order to address this limitation, we conducted three studies. In Study 1, we semantically sorted and factor analysed 1,772 personality items to identify a broad set of base facets. In Study 2, we conducted a systematic literature search to identify any potentially missing facets before writing items for all unique facets identified. In Study 3, we analysed and refined the resultant scales using confirmatory factor analysis, invariance analysis, and five assessments of 'discriminant validity evidence'. The ultimate result of these studies is a list of facets, and open access scales to assess them, that provide a first approximation to a comprehensive taxonomy of personality facets.

The importance of facets

Although each level of the personality hierarchy is interesting and useful (Mõttus et al., 2020), there are at least three compelling reasons to focus on facet-level traits. First, in existing models, facets represent the first factors identified above the item level and thus provide rich descriptions of stable and generalizable patterns of thought, feeling, and behaviour at a narrow-

trait level. In contrast, intermediate, broad, and super-ordinate factors offer multidimensional descriptions of human behaviour. In practice, however, the term facet is subject to different uses. Some scales described as facets are multidimensional, some are bloated specifics (i.e., little more than several slightly re-worded items; Cattell & Tsujioka, 1964), and often facets are assumed to be sub-components of a pre-specified set of broad factors (e.g., the Big Five). To clarify our use of the term, we see a facet as defined within the Cambridge dictionary, namely, "one part of a subject that has many parts", with the subject being human personality. We consider facets to be narrow, psychologically unidimensional (i.e., assessing only a single phenotype), yet generalizable (i.e., cross-situational) personality traits that are statistically identified as first-order factors, located directly above items.

Figure 1.

Illustrative example of a hierarchical model of personality based on the Big Five/Five-Factor Model tradition
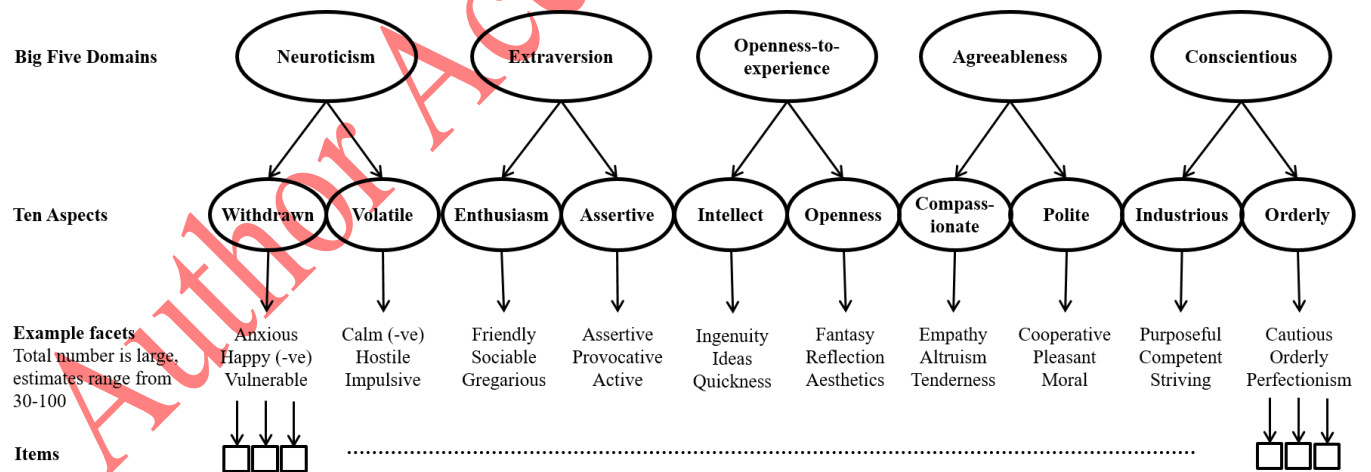
Second, compared with traits at higher-levels of the hierarchy, facets provide improved prediction of most important societal outcomes. Historically, the opposite has been argued, namely, that broad factors should be the most predictive because they contain the most important facet covariation, can be more reliable, and in most cases we wish to explain broad/complex

4

outcomes so we should also use broad/complex personality predictors (e.g., Ones & Viswesvaran, 1996; Salgado et al., 2015). However, the explanatory supremacy of facets, relative to broader factors, is now well established in most domains, including: work (Judge, al., 2013), education (Lounsbury, et al., 2003), online behaviour (Hughes et al., 2012), well-being (Anglim & Grant, 2014), attitudes (Soutter, & Mõttus, 2020), and more (e.g., Ashton, et al., 1995; Paunonen & Ashton, 2001). In reviewing facet vs. factor-level prediction within the organizational domain, Hughes and Batey (2017) stated, "to maximise predictive validity, narrow facets rather than broad factors – which lead to underestimates and/or distorted estimates of relationships – are evidently of greater value" (p. 164) but went on to note that "there is likely one large hurdle to overcome before targeted facet-level programmes are widely adopted in research and practice, namely, identifying a satisfactory list of facets … Thus, one major goal has to be to develop such a list" (p. 166).

Third, empirical estimates of intermediate, broad, and superordinate factors are typically calculated by combining facets. If important facets are missing then estimates of the number and nature of intermediate, broad or superordinate factors identified through techniques such as exploratory factor analysis are likely to be incorrect or incomplete (Booth & Murray, 2018; Condon et al., 2020; Mõttus et al., 2020; Saucier, Iurino, & Thalmayer, 2020, de Winter & Dodou, 2012).

Current facet taxonomies

The most widely researched personality trait taxonomies, the lexical Big-Five and the psychometric Five-Factor Model (FFM), posit five broad personality traits (Neuroticism, Extraversion, Openness-to-Experience/Intellect, Agreeableness, Conscientiousness) that subsume various narrower, facet-level traits, typically in the range of 15-30. In many cases, facets are chosen, at least in part, for pragmatic reasons (e.g., scale length). The same is true of

other broad taxonomic models such as the six-factor HEXACO model (Ashton & Lee, 2007, 2020; Lee & Ashton, 2004), AB5C (Hofstee, et al., 1992) and the 16PF (Conn & Riecke, 1994). Numerous studies have identified personality traits omitted by these taxonomies (e.g., Ashton, et al., 2000; DeRaad & Barelds, 2008; Jackson, et al., 1996; Lee, et al., 2000; Lee & Ashton, 2004; Saucier & Iurino, 2020). For example, Paunonen and colleagues (Paunonen & Jackson, 2000; Paunonen, 2002; Paunonen, et al., 2003) identified at least ten facets (e.g., Thriftiness, Humorousness, Integrity, Religiosity, Risk Taking) that were empirically distinct from the FFM and provided incremental explanation of important life outcomes (Paunonen et al., 2003).

Given that each taxonomy contains, to some extent, unique facets, we can conclude that no single list contains all facets of known importance. This is not necessarily problematic if we can combine facet lists. Indeed, modern studies that seek to map narrow personality traits tend to use facets from multiple lists (e.g., Saucier et al., 2020). For example, DeYoung et al. (2007) examined 'aspect' level traits, located between broad factors and narrow facets (see Figure 1) using the facets from the NEO-PI-R (Costa & McCrae, 1992) and the IPIP AB5C (Goldberg, 1999). However, combining facet lists in this way can be problematic for three main reasons: (i) many existing 'facet-level' scales are multidimensional (McCrae, 2015), (ii) facets from different scales with the same labels often have different definitions and item content (Jingle); and (iii) facets with different labels sometimes share definitions and item content (Jangle). One example of this Jingle-Jangle Fallacy (Ziegler et al., 2013) within DeYoung et al.'s (2007) analysis concerns the AB5C Calmness facet and the NEO Angry hostility facet, which have different names but very similar item content. Thus, in this case, and a number of others, the same items and/or facets are included twice, or more, distorting the final taxonomies.

6

Examining personality scales within the International Personality Item Pool (IPIP: Goldberg, 1999; Goldberg et al., 2006) provides stark illustration of the magnitude of multidimensionality, Jingle-Jangle, and construct proliferation within personality trait assessment (see also Schwaba et al., 2020). The IPIP contains over 3,000 open-source items and more than 250 scales that replicate previously developed inventories. The IPIP is an exceptional resource, without which many important studies of personality would not have been possible, including the current study. However, because the IPIP replicates existing scales it suffers from their problems. For example, at the time of writing, one can find six different scales labelled Anxiety, each with varying item content (Jingle). One can also find six scales with varying labels (e.g., dutifulness, rule-consciousness, prudence), where the item content predominantly focusses on rule following. Table 1 contains additional illustrations.

Table 1.

An Illustration of Jingle-Jangle within the IPIP

Jingle		Jangle	
IPIP HPI Calmness	IPIP CPI Calmness	IPIP NEO Anger	IPIP HPI Empathy
Rarely get irritated	Look for hidden meanings in things	Rarely get irritated	Rarely get irritated
Seldom get mad	Blurt out whatever comes into my mind	Am not easily annoyed	Am not easily annoyed
Am not easily annoyed	Believe in human goodness	Get upset easily	Get upset easily
Rarely complain	Am relaxed most of the time	Get irritated easily	Get irritated easily
Rarely lose my composure	Am not easily frustrated	Rarely complain	Rarely complain
Get upset easily	Act quickly without thinking	Get angry easily	Get annoyed with others' behaviours
Get angry easily	Expect things to fail	Lose my temper	Grumble about things
Get irritated easily	Have frequent mood swings		
Lose my temper	React intensely		

Snap at people	Judge myself more harshly than others do

To avoid the limitations of construct proliferation and Jingle-Jangle, it is necessary to treat scale labels with scepticism and ensure that item-banks remove duplicates but retain all important phenotypes. Two relatively recent studies have embraced such an approach.

First, the SAPA Personality Inventory (SPI, Condon, 2017; Condon et al., 2017) used 696 unique IPIP items and has currently collected data on over 200,000 participants. The SAPA methodology presents participants with random samples of items taken from a larger pool. To date, analyses have been reported for exploratory (N = 23,679, 13.9% mean response rate per item pair), replication (N = 54,855, 13.7% mean response rate per item pair), and confirmatory (N = 48,350, 13.6%, mean response rate) samples. Condon (2017) identified 27 facets which he suspects is an underestimate. Simulation evidence supports this conclusion given that unrestricted exploratory factor analyses with large and diverse item banks tend to fail to recover important factors (de Winter et al., 2009; MacCallum et al., 1999, 2001). Condon (2017, p.74) speculates that there may be as many as 100+ facets, but currently has a preferred estimate of 40-50.

Second, Rouco et al. (2022) administered 528 IPIP items selected based on ratings of protypicality for each of the Big Five to 722 U.S.A. undergraduates. The sample was divided in two. The first sub-sample was used to conduct exploratory factor analysis (EFA), using maximum likelihood estimation and geomin rotation, on items sorted into each of the Big Five domains. Minimum average partial tests and parallel analysis were used to estimate the number of facets in each domain. The second sub-sample was used to apply Confirmatory Factor Analysis to the scales derived from the first subsample. Finally, 387 German speakers completed

the 38 emergent facets, comprised of 202 items. The facets were subject to individual CFAs, invariance analysis, and a full inventory Bi-factor Exploratory Structural Equation Model. The study identified the second largest number of personality facets to date. By grouping facets prior to EFA, Rouco et al. (2022) avoided many of the issues associated with analysing large and varied item sets. However, previous studies would suggest that using only items that were prototypical of the Big Five restricted the number and nature of identifiable facets (e.g., Condon, 2017; DeRaad & Barelds, 2008; Saucier & Iurino, 2020; Saucier et al., 2020).

Current Study

In the current paper, we describe three studies designed to produce a preliminary but substantial step towards a comprehensive taxonomy of personality facets and develop a set of open source items to measure them. We adopt an iterative, multi-stage, multi-method approach. At the earlier stages of development, our goal was to be overly inclusive, such that we captured as much personality item content as possible. At the later stages, our goal was to refine, such that we eliminated redundancy and instances of jingle-jangle.

In Study 1, we build upon the strengths of Condon (2017) and Rouco et al. (2022). We compiled a database of 1,772 personality items from major inventories, sorted the items into semantically homogenous groups, retaining all items regardless of their protoypicality of any Big Few model, and factor analysed each group. Each facet identified using EFA was subsequently subjected to CFA and a panel review for theoretical coherence.

Despite the large number of items analysed in Study 1 it was possible that important personality traits were missed. Thus, in Study 2, we conducted a systematic qualitative review of the personality assessment literature to identify facets missing from the Study 1 list. In total, we

examined 706 personality facets at the item-level and added those that were unique. We then created open access items for each ostensibly unique facet identified.

In Study 3, the items were administered to a new sample and the psychometric properties of the facets were evaluated. We also conducted extensive tests of discriminant validity evidence to avoid redundancy.

Study 1

The goal of Study 1 was to identify as many potentially unique facets as possible from across the multiple inventories contained within the Eugene Springfield Community Sample (ESCS). Study 1 had five major steps. First, we sorted the facet scales and their 1,772 items into semantically homogenous groups. Simulation studies demonstrate that factor analysis with large and varied item banks is sub-optimal (de Winter et al., 2009; MacCallum et al., 1999, 2001). Grouping optimized the data to enhance accurate factor recovery by ensuring each item set had fewer factors, higher factor loadings, and more items per factor (de Winter et al., 2009; MacCallum et al., 1999, 2001). Semantic groups were created based on item content, rather than facet labels, in order to avoid problems of jingle-jangle and no item was removed due to a lack of prototypicality with the Big Five or any other Big Few model (Mottus et al., 2020). Second, we removed duplicate items to avoid bloated specifics that would undermine accurate factor identification. Third, we conducted exploratory factor analysis on each semantic group. Fourth, we examined each identified factor using confirmatory factor analysis. Fifth, we conducted additional analyses and conceptual reviews of the factors in order to assess their theoretical coherence and independence from other factors.

Method

Participants

10

Data for Study 1 was taken from the Eugene-Springfield Community Sample (ESCS: n=972). See Goldberg's (2008) technical report for detailed information on data collection and demographics. The first wave of data was gathered in 1993 from homeowners who agreed to complete questionnaires by post for at least five to ten years with 88% (n=757) of the original sample completing two of the last four questionnaires, demonstrating high levels of retention (Goldberg, 2008). All questionnaires were completed anonymously with identification numbers used to match responses. Participants received between $10 and $25 for each completed questionnaire (Grucza & Goldberg, 2007). In 1993, participants in the original sample ranged in age from 18 to 85 (M = 51).

Missing Data

We removed participants who had missing data for a complete personality questionnaire. Remaining missing responses were imputed using the Multivariate Imputation by Chained Equations (MICE) R package (van Buuren & Oudshoorn, 2021), generating 10 imputed data sets and randomly selecting a single dataset for analysis. Fewer than 5% of responses were imputed meaning that risk of bias was minimal (Little & Rubin, 2002; Schafer & Graham, 2002).

Measures

Seven omnibus personality inventories and four narrow scales were selected for analysis: The *NEO-PI-R* (Costa & McCrae, 1992; 240 items, assessing 30 facets, five-point Likert scale), *California Psychological Inventory* (CPI: Gough & Bradley, 2002; 462 items, assessing 13 scales, True/False response scale), *16 Personality Factor Questionnaire Fifth Edition* (16PF5: Conn & Rieke, 1994; 185 items, assessing 15 scales, three-point response scale), *Multidimensional Personality Questionnaire* (MPQ: Tellegen & Waller, 2008, 276 items, assessing 30 facets, True/False response scale), *Jackson Personality Inventory* (JPI-R: Jackson,

1994; 300 items, assessing 15 scales, True/False response scale), *HEXACO Personality Inventory* (Lee & Ashton, 2004, 192 items, assessing 24 facets, 5-point Likert scale), *Six Personality Factor Questionnaire* (Jackson, Paunonen, et al., 1996, 108 items, assessing 18 facets, five-point Likert scale), IPIP equivalent of the *Need for Cognition Scale* (Cacioppo & Petty, 1982; 10 items, five-point Likert scale), *Right-Wing Authoritarianism Scale* (RWA, Altemeyer, 1981; 14 items, nine-point Likert scale), *Social Dominance Orientation Scale* (SDO: Pratto, et al., 1994; 6-items, seven-point Likert scale), and IPIP *Machiavellianism Scale* (Social Astuteness, JPI: Jackson, 1994; 6 items, five-point Likert Scale). Further details, including year of administration, factor and facet names, and response formats are detailed in supplementary materials, S1.

Results and Discussion

Stage 1.1: Semantic grouping of items

Three of the authors (TB, DH, PI) created semantic item groupings with the goal of producing homogenous item sets. Authors read the items and facet definitions and iteratively developed groups in an inductive manner. Following the initial sort, the panel reviewed the groups, discussing any discrepancies until reaching a consensus. Review panels, with informed participants, provide an effective method for conceptual reviews such as this (Presser & Blair, 1994; Willis et al., 2000). In total, 1,772 items from 137 facets were sorted into 23 semantically homogenous groups. Table S1.1 within the supplemental material provides the scale labels and a general description of the item content within each semantic group.

Stage 1.2: Removal of redundant items

To identify identical or near identical items from within each semantic group, we first calculated tetrachoric, polychoric, or Pearson correlations, dependent on item response format,

for all items within each group, using `polycor` (Fox & Dusa, 2022) in R 2.13.0. We then ranked the correlations and examined any item pair with a correlation > +/- .40. The cut-off was arbitrary but deemed small enough to identify all highly similar item pairs. If two items had identical wording or were semantically synonymous, we removed one, retaining items that better corresponded to item writing guidelines (e.g., Irwing & Hughes, 2018) and those with a greater number of response options. We also removed one item if two items semantically varied but represented situation-general vs. situation-specific instantiations of the same underlying trait. For example, the items "I get lonely when I am left by myself" and "I dislike eating alone" correlated at .615. These items clearly have different content but both assess a preference for not being alone. The first item refers to a general or characteristic pattern that is applicable to many people and circumstances. In contrast, the second item is so specific that it would not apply to everyone (e.g., those with disabilities, dependents, large families) or every situation (e.g., schools, workplaces). Such items should be avoided when assessing personality (Stewart et al., 2022).

In total, we removed 355 items (20.03%) retaining 1,417 items, one of the largest item sets ever analysed to identify personality facets. Table S1.2, within the supplemental material, details all item pairs analysed, their correlation, and the removed items. Table S1.3 contains the summary statistics for the items removed from each inventory.

Stage 1.3: Exploratory factor analyses

Items from each of the 23 semantic groupings were subject to exploratory factor analyses in M*plus* 6.0 using weighted least squares means and variances estimation (WLSMV, Muthen, et al., 1997) and Geomin oblique rotations. We used multiple criteria to identify a plausible range of factors for each semantic group: parallel analysis (Horn, 1965), Velicer's (1976) Minimum Average Partial test, and the Very Simple Structure criterion (Revelle & Rocklin, 1979). The

application of these tests provided an indication of the number of factors but not precise estimates. Thus, we also explored factors +/- one at either end of the suggested range. For example, if the three tests suggested 4-8 factors, we examined solutions ranging from 3-9 factors. In identifying the optimal solution, we also considered the following criteria (Costello & Osbourne, 2005):

1. Each factor should be identified by at least three primary loadings greater than .30.

2. The solution should contain no Heywood cases.

3. The magnitude of the overall variance explained by the solution.

4. The number of items which failed to load on any factor, along with the number of items with substantial (>.30) cross loadings.

5. The coherence of the item content within each solution.

   Once we had established the optimal number of factors, solutions were refined by applying the following criteria;

1. Items with no loadings ≥.30 were removed, under the assumption that the latent variable should explain more than 9% of the common variance in the item.

2. Items with cross-factor loadings of approximately equal magnitude to the primary loading were removed. Such items were deemed to poorly differentiate between factors.

3. A final check for substantive factor coherence was conducted.

   In total, 225 factor solutions were estimated across the 23 item groupings. Table 2 provides a summary of the factor analyses. Table S1.4, within the supplementary materials, explains the rationale for the selection of the preferred solution. In total, 121 factors comprised of 969 items (i.e. 54.68% of the original item pool) were retained.

Table 2

Summary Results of the Exploratory Factor Analyses for the 23 Semantic Groupings

Group	Initial No. Items	Sample Size	PA	MAP	VSS	Range of Solutions Explored	Final. No. Factors	Variance Explained	Final No. Items
1	61	500	11	10	4	3-12	9	53.64%	46
2	52	467	9	6	3	2-10	6	47.01%	38
3	86	493	14	6	4	3-15	6	45.35%	51
4	60	560	11	5	5	4-12	5	42.36%	31
5	73	493	9	7	3	2-10	6	47.92%	51
6	77	561	14	6	4	3-15	7	39.22%	51
7	58	468	10	5	3	2-11	5	50.01%	35
8	81	550	18	7	5	4-19	6	41.00%	53
9	45	506	6	4	3	2-7	5	48.82%	29
10	44	572	6	4	4	3-7	4	42.42%	23
11	35	572	4	3	3	2-5	3	40.88%	24
12	20	572	4	2	2	1-5	2	35.97%	12
13	99	553	20	4	7	3-21	4	35.70%	58
14	54	553	9	3	3	2-10	3	40.67%	46
15	64	553	14	5	5	4-15	5	41.20%	51
16	47	572	12	5	6	4-13	5	40.86%	45
17	65	501	7	5	3	2-8	5	44.18%	28
18	72	485	16	4	4	3-17	5	40.24%	52
19	31	485	5	4	3	2-6	5	52.00%	26
20	73	554	13	5	4	3-14	7	43.73%	49
21	57	505	9	1	2	1-10	6	50.15%	48
22	83	525	12	7	5	4-13	6	40.38%	51
23	50	494	7	5	3	2-8	6	50.26%	41

Note: For thoroughness, we explored factors +/- one of either end of ranges suggested by tests of parallel analysis (PA), Minimum Average Partial (MAP), and Very Simple Structure (VSS) criterion the suggested range.

Stage 1.4: Confirmatory Factor Analyses (CFA)

Next, each of the 121 factors was subject to CFA to facilitate the identification of essentially unidimensional scales. Unidimensionality describes "whether the inferred manifest test distribution can be represented as a unidimensional, locally independent, monotone latent trait" (Stout, 2002, p. 487). Essentially unidimensional scales are not exactly unidimensional. Items always capture multiple sources of variance (e.g., substantive and measurement error, Ziegler & Hagemann, 2015). However, a well-fitting single factor CFA model with strong

loadings and no substantial correlated errors provides initial support for unidimensionality because the factor accounts for a substantial proportion of variance in the items, and once the variance due to the latent trait is removed, the items are unrelated (Hattie, 1984; 1985; Lord & Novick, 1968). To determine fit we used standard criteria derived from the simulation work of Hu and Bentler (1998, 1998): Root Mean Square Error of Approximation (RMSEA) ≤ 0.06, Standardized Root Mean Square Residual (SRMSR) ≤ 0.08, Tucker-Lewis Index (TLI) ≥ 0.95, and Comparative Fit Index (CFI) ≥ 0.95.

Factors that met the criteria for essential unidimensionality were retained without amendment. Factors that did not fit the criteria for essential unidimensionality were subjected to further analysis. In some cases, we made minor modifications, using modification indices to identify and remove problematic items before re-estimating the CFA models. In other cases, notably where scale items appeared multi-dimensional, we subjected the items to EFA, using the same process as described above, before testing the subsequent factors using CFA. In order to evaluate the improvement in fit between a single and multi-factor solution, we used the Satorra-Bentler chi-square difference test. We retained factors from models that showed improved fit and met the criteria for satisfactory solutions (e.g., 3+ items per factor etc.). For example, the initial factor 10 solution did not fit (CFI = .88; TLI = .86; RMSEA = .07), so was subjected to EFA which identified two related but distinct sub-factors concerning tendencies to be critical (e.g., If people are doing something wrong, I usually tell them what I think) and assertive (I am dominant, forceful and assertive). The subsequent two-factor model fit the data well (CFI = .98; TLI = .97; RMSEA = .04). Similarly, initial factor 75 did not fit (CFI = .89; TLI = .88; RMSEA = .04) and was subsequently split into two factors (CFI = .96; TLI = .95; RMSEA = .03) reflecting tendencies toward bad temper/frustration (e.g., Sometimes I feel like smashing things)

and laziness (e.g., I must admit that I often do as little work as I can get by with). Table S1.5, in the supplemental material, contains details of all revised factors.

In total, evaluating the factors using CFA produced 136 factors with a mean of 6.9 items (supplemental material, Table S1.5). Close fit was indicated in 133 cases by the CFI, in 110 cases by the TLI, and in 125 cases by the RMSEA. At least two fit indices supported close fit for all 136 factors.

Stage 1.5: Panel review of factors

At this stage, we had identified 136 factors that were potential candidates for inclusion within the personality facet taxonomy, using purely statistical methods. Factor analyses identify patterns of covariance that might represent meaningful underlying variables but also commonly identify spurious factors. Thus, some of the 136 factors might be statistically but not conceptually satisfactory. There are also two reasons to suspect that the 136 factors contained some duplication and redundancy. First, items that were highly similar, but which had varied response formats (e.g., binary vs. 5-point likert-type scales), sometimes formed separate factors, seemingly because the different response scales inflated within scale correlations. Second, semantic item groups were necessary to circumvent limitations of EFA with large and varied item banks (de Winter et al., 2009; MacCallum et al., 1999, 2001). However, it is possible that duplicate factors were derived within different semantic groups. Thus, we next conducted an iterative series of statistical-theoretical reviews to identify and remove redundant and conceptually weak scales.

In the first review, we calculated and ranked disattenuated factor inter-correlations in R 2.13.0, using Cronbach alpha as the estimate of reliability. Next, each panel member (TB, DH, PI) individually considered each pair of facets for redundancy in the rank order of their

disattenuated correlations. This process continued until 10 facet pairs in a row were considered unique. In the case of a facet pair adjudged to be duplicates, the panel members retained their preferred factor based on item quality as well as conceptual and psychometric robustness. For example, factors 17 (e.g., It is easy for me to strike up a conversation with someone) and 41 (e.g., Starting conversations with strangers: a. never gives me any trouble, b.?, c. is hard for me) correlated at .94 and contained highly similar content. Thus, we retained 17 and removed 41. During this process, panel members also identified scales that were evidently incoherent. Factor 83, for example, contained a variety of weak loading items with little obvious commonality (e.g., People should adapt their ideas and behavior to the group that happens to be with them at the time; I can tell as soon as I meet someone whether I will like that person or not; The person who provides temptation by leaving valuable property unprotected is about as much to blame for its theft as the one who steals it). This first panel review led to the removal of 56 factors and the combining of two pairs (16 & 42, 45 & 130) using CFA producing factors 137 and 138 respectively. Supplementary Table S1.6 contains a brief rationale for all judgements. In total, 78 facets were retained (see Table S1.7, supplemental material).

In a second review, a different panel (DH, PI, AT) systematically compared the item content of each facet with all other facets, because it was possible that some highly similar scales did not share large correlations, due to differences in response formats, and thus were missed in the first panel review. To aid this review, the panel developed working definitions for each facet aided by consultation of the Oxford English Dictionary (2015) and the APA Dictionary of Psychology (VandenBos, 2007). In doing so, the panel also identified several scales that lacked conceptual coherence. For example, Facet 46 was removed because it contained items assessing diverse phenomena (e.g., I find it hard to keep my mind on a task or job; I have reason for

18

feeling jealous of one or more members of my family; There seems to be a lump in my throat much of the time). In total, the second panel review adjudged six facets multidimensional, seven facet pairs synonymous, and four scales to be bloated specifics, arising due to method artefacts. Details of these judgments are contained in supplementary materials, Table S1.6. In total, the final list from Study 1 included 61 facets (see Table 3, Study 2), more than double that of most commonly used facet lists. We refer to this list as the 61 base facets.

Study 1 summary

Study 1 contributed to the overarching goal of this research by identifying a base-set of candidate facets to serve as a solid starting point for additional development. Sorting facets and items according to conceptual and semantic content, rather than any particular scale's adopted hierarchical structure, enabled facet identification to be based purely on empirical relations between personality items, in a manner that was unconstrained by the idiosyncrasies of existing measures (e.g., jingle-jangle, pragmatic facet selection, multidimensional facets, etc.). Further, because we assessed a large and varied set of personality items, from multiple major omnibus personality inventories, we are confident that the 61 facets cover much of the most important content in our field.

Confidence that the 61 base facets provide a sensible starting point for additional development is enhanced when they are compared to other facet lists. For example, we identified many more facet scales than previous studies that have adopted similar methods (e.g., Condon, 2017; Rouco et al., 2022; Watson et al., 2017) and studies that have used cluster analysis to assess large numbers of person-descriptive adjectives (e.g., Wood, Nye, & Saucier, 2010, who identified around 50 personality-like clusters and around 10 non-personality clusters, e.g., well-handicapped, slim-chubby). Similarly, Saucier and colleagues (2020) applied cluster analysis to

scale scores of 257 approximately facet-level scales and identified 62 clusters (plus 4 single item markers), many of which bear resemblance to the facets identified here.

Despite the large number of items analysed and the identification of a large number of psychometrically sound candidate facet scales, the biggest risk at this stage was that we had omitted important personality traits, either because insufficient items were included, semantic sorting separated items that might have coalesced, or the review panels made incorrect decisions. Thus, we considered the 61 base facets as exactly that, a base from which to build in subsequent studies.

Study 2

The major goal of Study 2 was to search the literature for personality facets beyond the 61 base facets. In addition, because many items comprising the base and additional facets were copyrighted and/or used varied response scales, we also wrote new items to ensure the final facet list comprised standardised and open access scales for unrestricted use by the research community.

Method

We searched for personality models via Google Scholar, Scopus, and Web of Science for the period 1960 to 2017 using the keywords: 'personality scale', 'personality facet', 'personality model', 'personality taxonomy', 'personality classification', 'personality repository', 'personality factor structure', and 'hierarchical model of personality'. We applied four inclusion criteria to each personality model or scale. First, the facet had to conform to the consensus definition of personality: patterns of thought, feeling, and behaviour (e.g. Costa & MacCrae, 2008; DeYoung, 2015; Fleeson, 2001; Tellegen, 1991). Personality was distinguished from some closely related constructs such as characteristic adaptations (DeYoung, 2015; McAdams & Pals,

2006), interests, and attitudes. Clinical facets of personality were included. However, we omitted non-personality clinical scales that assessed states or extreme symptomatology (e.g., the Dimensional Assessment of Personality Pathology – Basic Questionnaire Self-Harm scale). Second, facet scales were only included if they showed adequate psychometric properties. Third, we preferred personality scales that demonstrated theoretically coherent factor structures, even if developed post hoc. Fourth, we only included scales with open access item content, since otherwise it would have been impossible to evaluate the uniqueness of their content with respect to the 61 base facets.

Results and Discussion

Facets beyond the 61 base facets

The literature search located 706 facets, of which, 544 satisfied the four inclusion criteria. The supplementary materials lists all scales, separated into 398 scales from the IPIP (Table S2.1) and 146 from elsewhere (Table S2.2), as well as the 162 scales that did not meet our inclusion criteria (Table S2.3). The final facet list was produced by comparing the 544 scales to the 61 base facets (DH, PI, AT).

In total, 306 scales shared most or all of their item content with one of the baseline facets (Table S2.4, supplemental material). The fact that we identified approximately five instantiations of each of the 61 base facets suggests that they represent common, recurrent, and thus important personality traits in models developed from many different traditions. In short, the process lent notable credibility to the 61 base facets.

A further 217 scales were identified as multidimensional (Table S2.5 supplemental material), and thus shared item content with multiple base facets. Of these, 196 had sub-sets of items that were highly similar to those within the 61 base facets. Of the remaining 21 scales,

21

eight were synonymous with each other (e.g., IPIP 16PF Warmth, IPIP AB5C Warmth, ZKA-PQ

Social Warmth) leaving 16 scales deemed to provide content that was distinct from the 61 base

facets (Table S2.6, supplemental material). Thus, the final list contained 77 facets (see Table 3).

Table 3

Working definitions and labels for the 61 base facets identified in Study 1 and 16 unique facets

(*) identified in Study 2

Facet	Definition
Absorption	Propensity to feel engrossed in activities
Abstract thinking	Propensity to explore and discuss abstract ideas
Achievement Striving	Propensity to be ambitious and goal-oriented
Aesthetics	Enjoyment of artistic and aesthetic activities
Affability	Propensity to get along with others
Aggression	Propensity for hostile and threatening behaviour
Altruism	Propensity to enjoy helping others in an unselfish manner
Anger	Propensity to lose one's temper when frustrated
Anxiety	Propensity to feel apprehensive
Assertive	Propensity to behave in a self-assured and confident manner
Attention-seeking*	Propensity to draw attention to oneself and enjoy it
Callous	Propensity to be insensitive and indifferent towards others
Courageous	Propensity to be undeterred by danger, pain, or fear
Critical	Propensity to express adverse or disapproving comments or judgements
Dependability	Propensity to act in a reliable and responsible manner
Depression	Propensity to feel extreme negative affect
Detail Conscious	Propensity to pay careful attention to details
Dishonest-Opportunism	Propensity to cheat and act dishonestly to gain an advantage
Distractibility*	Propensity to be easily diverted from matters at hand
Eccentricity	Propensity to exhibit unconventional beliefs, thoughts and behaviours
Emotion-based decision making*	Propensity to make decisions based on feelings rather than logical arguments
Empathy	Propensity to try to understand and vicariously experience other's problems
Envy*	Propensity to feel resentful and discontented by others' wealth, qualities, or luck
Extrospection*	Propensity to examine others' thoughts, feelings, motives, and behaviour
Fair*	Propensity to treat others equally and impartially
Fantasy	Propensity to fantasise and day dream
Forgiving	Propensity to put aside feelings of resentment
Grandiosity	Propensity to exaggerate one's importance or ability
Gratitude*	Propensity to be thankful and grateful
Harm Avoidance	Propensity to avoid activities that are dangerous
Hedonism	Propensity to seek pleasure/fun
Honesty	Propensity to be truthful and act with integrity
Humour	Propensity to perceive or express the amusing aspects of a situation
Imagination	Propensity to generate ideas in the absence of direct sensory data
Impetuous	Propensity to act on the spur of the moment
Indecisive*	Propensity to struggle to make decisions quickly and efficiently
Inferiority	Propensity to feel inadequate and incapable compared with others
Insecure Attachment*	Propensity to fear and worry about being or becoming alone
Intellectually Curiosity	The desire to acquire a broad range of information

Intolerance	Propensity to reject views, beliefs, or behaviours that differ from one's own
Introspection*	Propensity to examine one's thoughts, feelings, motives, and behaviour
Leadership	Propensity to direct, co-ordinate, and organize others
Manipulative	Propensity to try to control and influence people or situations via devious methods
Need for Cognition	Enjoyment of extensive cognitive activity
Need for Social Acceptance	Propensity to seek out positive appraisal and acceptance from others
Novelty Seeking	Propensity to seek out novel experiences
Orderly	Propensity to be neat and tidy
Perseverance	Propensity to continue with and finish a task despite obstacles
Personal Disclosure	Propensity to share personal information
Planful	Propensity to plan
Positivity	Propensity to enjoy and look forward to life
Procrastination	Propensity to postpone and delay the beginning of a task
Provocative*	Propensity to try to cause anger and annoyance
Punitive	Propensity enforce discipline via punishment
Risk-aversion	Propensity to avoid activities or behaviors that entail danger, chance, or risk of loss
Rude*	Propensity to behave discourteously and lack manners
Rumination	Propensity to engage in negative repetitive thoughts
Self-control	Propensity to restrain impulses
Self-Efficacy	Propensity to hold the subjective perception that one is capable of performing
Self-Reliance	Propensity to rely on one's own resources
Sensation Seeking	The tendency to seek and enjoy thrilling and exciting activities
Sensitivity	Propensity to have intense responses to emotional stimuli
Sensitivity to Criticism	Propensity to respond negatively to criticism and teasing
Sociability	Propensity to enjoy the company of others
Social Boldness	Propensity to initiate social interactions
Social Confidence	Propensity to feel confident in social situations
Social Dependence	Propensity to seek out other's support during difficult times
Spirituality*	Propensity to believe in supernatural or universal powers
Stubborn	Propensity to adhere to rigid opinions
Surgency	Propensity to experience positive affect
Suspicious	Propensity to be apprehensive and mistrusting of others
Tolerance for Ambiguity*	Propensity to be comfortable with and enjoy ambiguous, unclear or uncertain situations
Traditionalism	Propensity to oppose change and maintain tradition
Vengeful	Propensity to retaliate and seek revenge
Vigour	Propensity to exhibit physical and mental energy
Warmth*	Propensity to be affectionate and kind
Worry	Propensity to feel mental distress or agitation due to concern about impending or anticipated events

Note: * = facets identified in Study 2

Developing open-access items for the 77 facets

We adapted Irwing and Hughes' (2018) recommendations on scale development to the current context, following a four-stage procedure of item development: i) construction of facet definitions; ii) item generation; iii) item review; and iv) item piloting.

We developed six items for each of the 77 facets because based on domain-sampling theory (e.g. DeVellis, 2016; Hinkin, 1998, Nunnally & Bernstein, 1994) it was considered that six items would: (i) allow adequate sampling of the range of phenotypes which represent a uni-dimensional facet, and (ii) should be sufficient to provide conceptual differentiation between facets (Lawson & Robins, 2021). Six-item scales also have statistical advantages, including: overidentifying CFA models (Brown, 2006a, b) and adequate information to estimate internal consistency.

We generated construct definitions (Table 3) before identifying or writing items. For all scales we replaced copyrighted items with IPIP items if they were clearly equivalent, otherwise, we generated new items or modified the wording. New items were generated following the item writing guidelines recommended by Irwing and Hughes (2018, p.17), and all items were subjected to panel review, which is generally found to be an effective procedure (Demario & Landreth, 2004; Presser & Blair, 1994). The result of Study 2 was 462 items developed to assess 77 facets. Of those, 210 items were written by the authors, 208 originated from the IPIP, and 44 were derived from freely available research scales (Table S3.1, supplemental material, contains all items).

Study 3

The goal of Study 3 was to assess the psychometric properties of the 77 newly developed facet scales regarding three major concerns that are crucial in scale development. First, we

examined whether the newly developed scales evidenced strong psychometric properties. We examined CFA fit, factor loadings, average variance extracted (AVE: Fornell & Larker, 1981), and MacDonald's Omega (Revelle & Condon, 2018; Revelle & Zinbarg, 2009).

Second, we examined whether the new scales were essentially unidimensional (Hattie, 1985, Stout, 1987). Unidimensionality is crucial for theory development and empirical conclusions because although items capture multiple sources of variance, essentially unidimensional scales provide strong assessment of a single construct, which enables unambiguous interpretation (Hattie, 1985; Stout, 1987). Confirmatory factor analysis is a useful tool for initial assessment of unidimensionality because it enables precise modelling of a latent variable, errors of measurement, correlated errors between items, and tests of the fit of this model (Bollen, 1989, Hattie, 1985).

Third, we examined whether each facet was unique (i.e., 'discriminant validity'; Campbell & Fiske, 1959). We adopted a multi-faceted approach to assessing discriminant validity evidence, influenced by a range of recent theoretical contributions (e.g., Hughes, 2018; Lawson & Robins, 2021) and simulation studies (e.g., Rõnkko & Cho, 2020; Shaffer, et al., 2016). All of these guiding frameworks argue against strict adherence to any particular statistical framework and recommend multi-stage evaluations to assess whether two scales are related but distinct or redundant. Accordingly, we considered evidence from four quantitative procedures (Rõnkko & Cho, 2020; Shaffer, et al., 2016): dissatenuated correlation coefficients using corrections based on (1) internal consistency estimates and (2) SEM; the absolute and relative fit of one- and two-factor models (3); and the relative magnitude of the dissatenuated correlation and the square root of the AVE (4). In addition, we also used conceptual criteria to aid our judgements (Hughes, 2018; Lawson & Robins, 2021). It is possible that scale pairs might appear

distinct based on quantitative evidence but share items with equivalent meaning or appear

redundant based on quantitative evidence whilst having clearly distinct conceptual definitions

and item content (Hughes, 2018; Lawson & Robins, 2021; Rosenbusch et al., 2019). Thus, a

judgement of conceptual similarity, based on construct definitions and item overlap provided

important evidence that informed our examination of uniqueness and redundancy.

Method

Participants

A priori power analysis using the MBESS package in R version 4.3.0, testing for a

population RMSEA = .08, with a 90% confidence interval width of .04, suggested a minimum

sample size of 819 (Kelley & Lai, 2018). Given that, in total, 1,096 participants completed the 77

personality facets online, we had adequate power to test for the fit of each of the 6-item latent

variables. The sample aggregated data from three sources, collected in parallel, to enhance size

and demographic diversity. A European sample ($N = 341$) was collected via the German data

collection company Lime Survey. The company used a simple random sampling technique to

select respondents older than 18 from their sampling frame of European nationals. Participants

were paid €11.20 for completing the survey. A U.S.A sample ($N = 523$) was collected via

convenience sampling on Mechanical Turk (i.e., Mturk), a web portal developed by Amazon to

facilitate effective data collection by linking researchers to a large pool of primarily US-based

respondents (Goodman, et al., 2017). Participants were paid $7.56 (20% higher than the standard

fee for a survey of this size). A broader international sample ($N = 232$) was collected directly by

the research team employing convenience and snowball sampling to target an international (i.e.,

non E.U. or U.S.A) network of contacts, including colleagues, friends, and acquaintances.

Prospective participants were contacted via social media messages or e-mail and each was asked

to further advertise the survey among their own contacts. Participants were incentivised by the offer of a personalised personality report comparing their responses to those of the full sample.

The combined sample (N= 1,096) comprised 50.8% males, 49.1% females, and .1% transgender, with ages ranging from 18 to 75 years (M = 38.4; SD = 12.8). The age distribution was slightly skewed towards younger ages (*Skew = .739; Kurtosis = -.322*), with the 18–29 and 30–39 years age groups collectively representing 61.7 per cent of the sample. Respondents in the sample represented 53 different countries, with the highest proportion of respondents residing in the United States (43.3 per cent), the United Kingdom (13.9 per cent), Germany (6.5 per cent), and India (4.5 per cent) (see Table S3.1 for exact details). White respondents were over-represented (73.7 per cent) compared to other ethnicities (e.g., South Asians, 5.4 per cent; North African, .2 per cent). The sample was skewed towards higher levels of education, with 64.5% of respondents possessing undergraduate and postgraduate degrees. Respondents worked in a wide variety of occupations, the most prevalent including office and administrative positions (11.8%), education and training (10.7%), computers and mathematics (10.7%), sales (9.1%), business and financial operations (8.8%), management (6.3%), and arts, design, entertainment, sports, and media (5.5%).

Missing Data

Missing data was present across only nine cases (.82 per cent of the total sample). In all nine cases, the participants completed just a few items, with missingness ranging from 78 per cent to 91 per cent. Little's (1988) Missing Completely At Random test ($\chi^2(547) = 508.15$, *p = .89)* indicated that the missingness was truly random. Thus the nine cases were removed Listwise, since this method is unbiased with <1% missing data, and when missingness is

completely at random (Arbuckle, 1996, Wotke, 2000). The final sample numbered 1,087 respondents.

Measures

First, respondents read an introduction to familiarise themselves with the purpose of the study, including an information sheet and consent form. Next, participants completed demographic questions (age, gender, current country of residence, country of birth, ethnic origin, highest level of education, and occupation) before the 77 personality facets. Facets were randomised to minimise order effects (Couper, 2008) and counteract potential fatigue effects for scales presented towards the end of the questionnaire (Ackerman & Kanfer, 2009). All items had a seven-point Likert-type response scale ranging from 'Strongly Disagree' (1) to 'Strongly Agree' (7). The seven-point response scale was chosen because it provides greater variance and reliability than fewer response options (Cook, et al., 2001), with gains typically plateauing around six or seven response options (Finstad, 2010; Simms et al., 2019). Further, it provides a midpoint, neutral, response option suitable for those who feel truly neutral (Cox, 1980; Weems & Onwuegbuzie, 2001). One caution here is that when items are ambiguous, the neutral response option tends to reflect "it depends" response processes (Kulas, & Stachowski, 2009; 2013). Of course, the same problems arise when midpoint options are omitted but the error variance is spread throughout the scale. Thus, for this reason and many others (Hughes, 2018), it was important to ensure that items did not require clarification or contextualisation (Kulas, & Stachowski, 2009; 2013).

Ethics

The study received ethical approval from the University of Manchester in accordance with the ethical guidelines of the British Psychological Society (BPS, 2018).

Results and Discussion

Confirmatory Factor Analyses

The 77 facets were subject to individual CFAs using the weighted least squares means and variances (WLSMV) adjusted estimator in Mplus 7.4 (Muthén & Muthén, 1998-2015). WLSMV is suitable for ordinal-level data and is robust to violations of multivariate normality (Flora & Curan, 2004; Booth & Hughes, 2014). CFI and TLI values of $\geq$ .90 were taken as indicative of acceptable model fit (Hu & Bentler, 1998, 1999; Schermelleh-Engel, et al., 2003). Typically, RMSEA values of $\leq$ .08 are also taken as indicative of acceptable model fit. However, recent simulation studies suggest that the .08 criteria is inappropriate for scales with relatively few items and strong loadings (Kenny, et al., 2015; McNeish, et al., 2018). This so called 'reliability paradox' leads to inflated RMSEA values that 'punish' unidimensional and highly reliable scales (McNeish et al., 2018). In such cases, the RMSEA is a less useful indicator of fit, with values around .2 typically considered acceptable.

Psychometric Robustness and Essential Unidimensionality.

Of the 77 scales, 67 met the criteria for moderate or close fit according to the CFI, 45 according to the TLI, and 33 according to the RMSEA (Table S3.2). Following modest modifications (e.g., removal of 1 or 2 items, 40 items in total, just 8.66%) all 77 facets fit according to the CFI and TLI, and 55 according to the RMSEA. Items were removed based on the magnitude of factor loadings, the sum of their modification indices, and item clarity. Psychometric guidelines typically recommend testing at least twice as many items as required, so the relatively small number of items omitted from the final scales supports the current process of item and scale development (Kline, 1993).

The Average Variance Extracted (AVE) values provided further evidence of essential unidimensionality and factorial quality of the facets. The AVE values were > 50% for 66 scales,

> 40% for 9 scales, and >35% for 2 scales. Expressed in perhaps more interpretable form, average item loadings were large, exceeding .71 for 66 scales, .63 for 9 scales, and .60 for 2 scales. It is desirable for factor loadings to be as high as possible (Mulaik, 2018), provided this is not due to the presence of bloated specifics (Cattell & Tsujioka, 1964). Table 4 contains summary descriptive statistics, fit indices, reliabilities, and the AVEs for original and modified scales. Supplementary Table S3.2 contains complete details of the CFAs applied to the original 77 scales and Table S3.3 describes the rationale for each scale modification.

Table 4

Summary of the psychometric properties of the 77 original and modified scales

Facet	# of items	*M*	*SD*	Skew	Kurt-osis	CFI	TLI	RMSEA	W_t	*a*	AVE (%)
Absorption	6	4.29	1.26	-.32	-.28	.940	.900	.238 [.221, .255]	.85	.85	55.33
Abstract Thinking	6	5.00	1.31	-.42	-.44	.984	.973	.157 [.140, .174]	.90	.89	66.15
Achievement-striving	6	5.29	1.17	-.55	-.15	.968	.947	.298 [.281, .315]	.92	.92	76.12
Aesthetics	6	5.31	1.02	-.66	.75	.983	.972	.098 [.081, .115]	.81	.81	50.73
Affability	6	4.66	1.38	-.45	-.31	.968	.946	.259 [.243, .276]	.92	.92	72.63
Aggression	6	2.64	1.25	.70	-.04	.975	.958	.160 [.144, .177]	.86	.85	61.63
Altruism	6	5.02	1.02	-.52	.61	.976	.960	.134 [.117, .151]	.83	.82	53.07
Anger	5	3.33	1.31	.29	-.49	.986	.973	.155 [.133, .178]	.86	.85	61.18
Original model	6	3.40	1.22	,16	-.30	.887	.812	.332 [.316, .349]	.84	.83	55.51
Anxiety	6	3.45	1.58	.28	-.82	.996	.993	.118 [.101, .135]	.94	.94	77.20
Assertive	6	4.76	1.04	-.50	.50	.969	.948	.126 [.109, .143]	.81	.80	46.97
Attention-seeking	5	5.10	1.17	-.63	.41	.964	.929	.224 [.202, .247]	.84	.83	58.34
Original model	6	5.02	1.17	-.58	.40	.886	.809	.301 [.285, .318]	.82	.80	51.18
Callous	6	2.27	1.21	1.21	1.22	.995	.991	.124 [.107, .141]	.93	.93	76.67
Courageous	5	3.83	1.17	.19	-.19	.960	.919	.186 [.164, .209]	.79	.78	48.26
Original model	6	3.72	1.17	. 20	-.22	.902	.837	.251 [.235, .268]	.82	.82	50.23
Critical	5	3.26	1.16	.23	-.33	.994	.987	.087 [.065, .110]	.83	.82	54.72
Original model	6	3.26	1.10	.17	-.23	.928	.879	.230 [.213, .247]	.83	.83	51.81
Dependability	6	5.64	.95	-.86	.96	.997	.996	.064 [.047, .082]	.90	.90	67.37
Depression	6	2.89	1.58	.66	-.53	.995	.992	.128 [.112, .145]	.95	.95	80.07
Detail Conscious	5	5.32	.97	-.45	-.10	.987	.975	.112 [.090, .135]	.81	.81	52.06
Original model	6	5.31	.98	-.41	-.21	. 924	.873	.233 [.217, .250]	.81	.81	50.07
Dishonest-Opportunism	5	3.02	1.24	.30	-.42	.982	.964	.118 [.097, .142]	.78	.76	49.22
Original model	6	3.01	1.16	.22	-.34	.927	.879	.184 [.167, .201]	.77	.76	44.33
Distractibility	6	4.04	1.83	-.11	-1.10	.979	.965	.212 [.195, .229]	.92	.92	73.45
Eccentricity	5	3.81	1.34	-.09	-.50	.985	.971	.167 [.145, .190]	.88	.88	63.98
Original model	6	3.92	1.15	-.01	-.38	.984	.973	.134 [.118, .151]	. 83	.80	53.31
Emotion-based decision making	6	5.05	1.16	-.69	.40	.963	.938	.191 [.174, .208]	.86	.86	56.27
Empathy	5	4.83	1.11	-.41	.20	.992	.983	.089 [.067, .113]	.81	.81	53.10
Original model	6	4.85	1.07	-.29	.18	.926	.877	.206 [.189, .222]	.81	.80	49.22
Envy	6	3.15	1.29	.13	-.67	.982	.970	.192 [.192, .209]	.92	.92	71.87
Extrospection	6	3.05	1.40	.44	-.53	.969	.948	.211 [.194, .228]	.90	.90	66.60
Fair	6	5.61	.97	1.06	.48	.975	.958	.153 [.136, .170]	.85	.85	58.62
Fantasy	6	4.12	1.40	-.12	-.54	.962	.937	.277 [.261, .294]	.92	.91	71.08
Forgiving	5	4.56	1.22	-.55	.23	.993	.986	.103 [.081, .127]	.87	.86	60.66
Original model	6	4.52	1.22	-.49	.25	.923	.872	.278 [.262, .295]	.88	.88	57.28
Grandiosity	6	2.91	1.28	.37	-.66	.998	.997	.053 [.035, .071]	.89	.88	63.33
Gratitude	5	2.96	1.33	.58	-.12	.982	.964	.145 [.123, .168]	.72	.71	53.62
Original model	6	3.01	1.28	.53	-.05	.891	.819	.282 [.266, .299]	.74	.73	49.50
Harm Avoidance	6	5.57	1.23	-.78	-.-5	.974	.957	.156 [.139, .173]	.86	.86	61.48
Hedonism	4	4.08	1.21	-.03	-.36	.975	.925	.215 [.181, .252]	.80	.79	59.63

Revision 1	5	3.60	.93	-.18	-.06	.922	.844	.285 [.267, .307]	.81	.81	52.54
Original model	6	4.13	.95	-.25	.01	.917	.861	.224 [.207, .241]	.77	.74	43.90
Honesty	6	5.61	.97	-.83	.74	.962	.937	.204 [.188, .221]	.87	.86	61.48
Humour	5	4.95	1.09	-.59	.58	.992	.984	.094 [.072, .118]	.84	.84	56.66
Original model	6	4.81	.99	-.44	.47	.966	.943	.150 [.134, .167]	.81	.79	48.11
Imagination	4	4.47	0.76	-.19	.73	.998	.993	.096 [.062, .134]	.87	.85	62.4
Original model	6	4.81	1.11	-.61	.71	.968	.947	.187 [.171, .204]	.87	.86	57.35
Impetuous	6	3.29	1.34	.28	-.66	.998	.996	.090 [.073, .107]	.91	.91	69.25
Indecisive	6	3.39	1.37	.28	-.70	.985	.976	.156 [.140, .173]	.90	.89	64.85
Inferiority	6	3.08	1.36	.47	-.34	.987	.978	.137 [.129, .162]	.89	.89	64.67
Insecure Attachment	6	2.52	1.06	.74	.13	.950	.916	.250 [.234, .267]	.89	.88	66.02
Intellectual Curiosity	5	5.39	1.05	-.74	.61	.958	.916	.236 [.214, .259]	.84	.84	59.32
Original model	6	5.20	.97	-.50	.34	.957	.928	.182 [.165, .199]	.81	.79	50.38
Intolerance	4	3.02	1.07	.31	-.24	.993	.979	.063 [.030, .103]	.65	.64	36.88
Revision 1	5	3.09	.97	.21	-.09	.940	.880	.127 [.105, .150]	.65	.64	31.44
Original model	6	3.02	.93	.22	.08	.841	.734	.221 [.204, .237]	.71	.70	35.65
Introspection	6	2.86	1.31	.54	-.35	.963	.938	.218 [.188, .224]	.89	.89	65.15
Leadership	5	4.56	1.16	-.42	.12	.970	.940	.161 [.133, .169]	.82	.81	52.36
Original model	6	4.45	1.13	-.32	.15	.938	.896	.183 [.167, .200]	.83	.82	49.15
Manipulative	6	2.95	1.23	.49	-.14	.954	.924	.210 [.193, .227]	.86	.86	58.63
Need for Cognition	4	4.40	1.29	-.05	-.28	.980	.940	.170 [.136, .207]	.78	.77	48.68
Revision 1	5	4.44	1.15	-.06	.05	.872	.745	.310 [.288, .333]	.76	.75	44.54
Original model	6	4.45	1.13	-.13	.18	.811	.685	.303 [.287, .320]	.79	.79	45.97
Need for Social Acceptance	4	3.61	1.22	-.10	-.50	.978	.935	.170 [.136, .207]	.76	.75	48.68
Revision 1	5	3.74	1.11	-.15	-.20	.906	.812	.231 [.209, .254]	.74	.72	41.44
Original model	6	3.79	1.11	-.14	-.18	.827	.712	.322 [.306, .339]	.79	.78	45.91
Novelty Seeking	5	5.29	.99	-.52	.24	.954	.908	.182 [.160, .205]	.77	.77	48.66
Original model	6	5.08	.98	-.42	.21	.904	.840	211 [.195, .228]	.79	.79	46.00
Positivity	6	5.02	1.34	-.86	.31	.991	.984	.145 [.129, .162]	.92	.92	69.88
Orderly	5	4.87	1.16	-.29	-.20	.968	.936	.164 [.170, .206]	.84	.83	52.34
Original model	6	4.85	1.21	-.28	-.41	.924	.873	.260 [.244, .277]	.86	.85	57.03
Perseverance	5	5.31	1.15	-.65	.07	.988	.975	.220 [.198, .243]	.90	.90	72.14
Original model	6	5.29	1.15	-.55	-.13	.928	.880	.429 [.413, .446]	.90	.90	72.97
Personal Disclosure	5	5.86	.86	-.94	.89	.985	.970	.151 [.113, .149]	.87	.87	61.40
Original model	6	4.11	1.3	-.18	.45	.911	.851	.305 [.289, .322]	.88	.88	61.07
Planful	6	5.38	1.01	-.58	.17	.994	.990	.086 [.070, .104]	.88	.87	62.37
Procrastination	5	3.27	1.40	.16	-.80	.999	.998	.066 [.044, .091]	.90	.89	69.32
Original model	6	3.16	1.29	.11	-.72	.939	.898	.357 [.340, .374]	.88	.88	68.81
Provocative	6	2.23	1.23	-.25	-.49	.981	.969	.172 [.156, .189]	.91	.91	72.69
Punitive	5	4.08	1.35	-.23	-.55	.969	.938	.156 [.134, .179]	.81	.81	50.92
Original model	6	4.08	1.35	-.24	-.59	.917	.862	223 [.207, .240]	.83	.83	51.28
Risk-aversion	5	2.94	1.16	.35	-.34	.994	.987	.104 [.082, .127]	.84	.84	58.66
Original model	6	3.09	1.13	.19	-.34	.936	.893	.256 [.239, .273]	.85	.84	55.13
Rude	6	5.19	1.11	-.70	.73	.985	.975	.091 [.075, .109]	.80	.78	48.82
Rumination	6	4.00	1.48	-.16	-.69	.986	.976	.156 [.140, .173]	.91	.91	68.93
Self-control	5	4.70	1.07	-.18	-.22	.969	.938	.143 [.121, .166]	.77	.76	45.18
Original model	6	4.52	1.07	-.04	-.18	.883	.805	.230 [.213, .247]	.78	.78	43.23

		M	SD			CFI	TLI	RMSEA	W_t	a	AVE
Self-Efficacy	5	5.33	1.02	-.63	.34	.998	.995	.083 [.061, .107]	.90	.90	71.40
Original model	6	5.13	.95	-.33	.13	.997	.995	.069 [.053, .087]	.87	.85	60.56
Self-Reliance	6	4.98	1.30	-.59	.13	.981	.969	.217 [.201, .234]	.92	.92	73.50
Sensation Seeking	6	3.17	1.59	.32	-.90	.989	.982	.147 [.130, .164]	.92	.92	57.98
Sensitivity	4	4.42	1.20	-.25	-.19	.995	.985	.052 [.018, .092]	.64	.61	36.98
Original model	6	4.29	.99	-.21	.45	.860	.766	.139 [.122, .156]	.70	.68	25.23
Sensitivity to criticism	6	4.17	1.32	-.16	-.34	.948	.913	.258 [.242, .275]	.88	.88	61.83
Sociability	6	4.11	1.17	-.38	.22	.966	.944	.190 [.174, .207]	.85	.83	53.50
Social Boldness	4	3.85	1.48	-.18	-.75	.995	.986	.139 [.105, .176]	.87	.86	66.77
Revision 1	5	3.86	1.44	-.18	-.60	.928	.875	.351 [.329, .374]	.87	.87	65.90
Original model	6	3.84	1.38	-.19	-.45	.915	.859	.319 [.303, .336]	.88	.87	62.67
Social Confidence	6	4.19	1.38	-.27	-.36	.957	.929	.280 [.264, .297]	.88	.88	62.88
Social Dependence	6	3.80	1.30	-.10	-.43	.946	.910	.249 [.232, .266]	.88	.88	60.92
Spirituality	6	5.12	1.10	-.42	.18	.994	.990	.140 [.123, .157]	.93	.93	76.63
Stubborn	6	4.09	1.21	-.18	-.30	.958	.930	.186 [.169, .203]	.86	.86	55.87
Surgency	6	4.63	1.21	-.53	.25	.968	.947	.230 [.214, .247]	.88	.87	60.73
Suspicious	6	3.75	1.18	.00	-.06	.958	.930	.181 [.165, .198]	.86	.86	56.02
Tolerance for Ambiguity	4	3.74	1.09	.27	.37	.992	.977	.083 [.050, .122]	.71	.70	43.13
Revision 1	5	3.79	1.05	.18	.50	.947	.895	.144 [.108, .167]	.72	.69	37.18
Original model	6	3.82	1.05	.23	.55	.879	.798	.214 [.198, .231]	.77	.76	40.92
Traditionalism	6	4,54	1.07	-.28	-.11	.941	.902	.153 [.137, .170]	.78	.77	41.87
Vengeance	5	3.03	1.24	.26	-.62	.985	.970	.153 [.131, .176]	.84	.83	58.96
Original model	6	3.17	1.16	.10	-.50	.925	.875	.255 [.238, .272]	.83	.82	52.98
Vigour	5	4.22	1.25	-.29	-.33	.961	.922	.203 [.181, .226]	.83	.82	53.72
Original model	6	4.28	1.19	-.24	-.21	.876	.793	.286 [.269, .302]	.83	.82	50.00
Warmth	6	3.99	1.12	-.23	-.28	.984	.973	.194 [.178, .211]	.91	.91	70.80
Worry	6	4.25	1.34	-.17	-.43	.964	.940	.228 [.212, .245]	.87	.86	57.97

Note. Table S3.2 shows the standardised factor loadings. Table S3.3 describes the rationale for each item removal. *M* = mean; *SD*: Standard Deviation; CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; RMSEA: Root Mean Square Error of Approximation; W_t = Omega; *a* = alpha; AVE= Average Variance Extracted.

Investigating Potential Redundancy

Reliability corrected correlations of .85 (Clark & Watson, 1995; Kline, 2011) are typically argued to indicate potential redundancy. However, we adopted a more conservative approach that reduced the risk of missing redundant facets. Specifically, we examined scale pairs with disattenuated correlations ≥ .70, correcting them for unreliability using McDonald's Omega (Revelle, 2016, 2022). In total, we identified 32 pairs of highly correlated scales for more thorough examination using four statistical criteria (Shaffer et al. 2016, Rõnkkö & Cho, 2020) and a conceptual and semantic review of the items (Hughes, 2018; Lawson & Robins, 2021).

First, we compared two reliability corrected correlations (Omega corrected scale scores and CFA based latent correlations) against a cut-off of ≤|.85| (Clark & Watson, 1995; Kline, 2011). Second, we compared the upper 95% confidence interval of the CFA estimated correlation between against a cut-off of <|.80|, which was recently suggested as a preferred strategy by Rõnkko and Cho (2020). Third, we compared the fit of one-factor and two-factor CFA models (Rõnkkö & Cho, 2020; Shaffer et al., 2016). Fourth, we compared the square root of the AVE for both latent variables against the factor correlation. Fornell and Larcker (1981) suggested that if the factor correlation was higher than the square root of both AVEs, this suggests redundancy. Fifth, we conducted a conceptual review of the scale definitions and item content.

Table 5 contains summary data from the statistical criteria and supplementary Table S3.4 displays a visual summary of which criteria were met or failed. Only 2 of the 33 pairs (social confidence and social boldness, provocativeness and rudeness) shared a corrected correlation above .85. In all cases, the two-factor CFA fit better than did the one-factor CFA. Fourteen pairs shared a CFA-based correlation with an upper 95% confidence interval ≥.80. Sixteen pairs had a

single squared AVE smaller than the correlation, and in 4 cases, both square roots were smaller (Fornell & Larcker, 1981). Whenever the statistical criteria identified potential problems with regard to redundancy, the panel conducted a conceptual review which formed the final decision (Hughes, 2018; Lawson & Robins, 2021). The panel (DH, PI, AT) first examined whether the item content supported a conceptual distinction, whether the constructs were multi-dimensional, and whether the construct could be considered an outcome of more fundamental personality traits (Lawson & Robins, 2021). In 6 of the 32 cases it was decided that there was a substantive problem.

Social boldness and social confidence exhibited the highest dissattenuated correlation of .85, and failed statistical criteria 1, 3, 4 and 5 (Tables 5 and S3.4). There was a clear overlap in item content, which referred to feeling comfortable in social situations. In consequence, we merged these facets into a single scale retaining the label social confidence.

The positivity and surgency items assessed a propensity toward positive affect and cognition, in consequence these items were collapsed into a single scale labelled positivity. Statistical criteria 2 and 4 and the semantic assessment supported this decision (Tables 5 and S3.4).

The aesthetics and sensitivity items assessed a propensity to appreciate art and nature, so these items were combined into a scale labelled aesthetics. Again, this decision was supported by statistical criteria 2 and 4 (Tables 5 and S3.4).

The case of leadership and assertiveness was more complex. Criterion 4 suggested a problem, yet according to the semantic assessment, the leadership items assessed the ability to give directions and make decisions, whereas the assertiveness items assessed a propensity to express ones point of view in a forceful manner. Although there is a clear distinction between the

36

two, the question arose as to whether leadership was a fundamental personality trait or not. The panel decided that the leadership items assessed an outcome of personality rather than personality per se, with leadership most likely a multidimensional outcome of more fundamental traits. We therefore merged these scales into a single scale of assertiveness, retaining those items regarding forceful expression and removing those regarding the outcome/quality of leadership.

The panel considered the facets of rudeness, aggression, and provocativeness in parallel. All failed statistical criteria 2 and 4 and were highly correlated (Tables 5 and S3.4). Provocativeness and aggression contained item content describing different forms of aggression. The rudeness items were arguably highly culture-specific indicators of passive aggression. Thus, we discarded the rudeness scale and entered provocativeness and aggression into a single factor CFA, which fit, supporting the idea that the items assessed a single unidimensional facet of aggression. We then produced a final 6-item scale of aggression, which fit a CFA, by choosing the most diverse and high quality items.

At this point, we conducted a final qualitative review of all scales. This review identified that harm avoidance and risk-aversion were definitionally indistinguishable and shared highly similar item content. However, the two correlated at -.69, just below our .7 threshold, because they were negatively and positively worded measures of the same construct. An amalgamated 6-item scale showed close fit: (CFI = .986, TLI = .977, RMSEA = .121[.105, .139).

Following tests for redundancy, 70 psychometrically robust and discriminant facets remained. Supplementary Table S3.5 includes the item content, factor loadings, and CFA fit of the final 70 facet scales.

Having identified the final set of facets, we tested for scalar invariance across our three data sets (i.e., European, U.S.A, and International samples) using the Millsap and Tein (2004)

approach for categorical/ordinal items. Scalar invariance assesses cross-group and cross-source invariance and is a minimum standard for reliable cross-group scale score comparisons (Mellenbergh, 1997; Millsap & Kim, 2018). According to standard model fit cut-offs (i.e., two or more of the CFI, TLI ($\geq$ .95) and SRMR ($\leq$ .08) suggest close fit), all 70 facet scales met the criteria for scalar invariance.

Table 5

Single and two-factor model fit, latent correlations, and square root of AVE for pairs of latent variables with rank-ordered corrected

correlations greater than .70

Correlated Constructs		r_{ss}	Single-factor model			Two-factor model			r_l 95%CI	$\sqrt{AVE_1}$	$\sqrt{AVE_2}$
			CFI	TLI	RMSEA	CFI	TLI	RMSEA			
Perseverance	Self-Efficacy	.71	.924	.902	.253 [.245-.262]	.987	.982	.107 [.099-.116]	.72 [.70,.74]	.85	.85
Manipulative	Provocative	.71	.897	.874	.194 [.187-.201]	.957	.947	.126 [.119-.133]	.73 [.69,.77]	.77	.85
Affability	Warmth	.71	.902	.880	.238 [.231-.245]	.965	.957	.143 [.137-.150]	.73 [.69,.77]	.85	.84
Gratitude	Warmth	.71	.918	.898	.217 [.209-.225]	.976	.970	.118 [.111-.126]	.69 [.65,.73]	.73	.84
Rumination	Inferiority	.72	.901	.879	.210 [.203-.217]	.985	.981	.084 [.077-.091]	.73 [.69,.77]	.83	.80
Self-efficacy	Dependability	.72	.924	.905	.217 [.209-.225]	.991	.988	.077 [.069-.085]	.74 [.72,.76]	.85	.82
Callous	**Rude**	.72	.948	.937	.167 [.160-.174]	.991	.989	.068 [.061-.076]	.75 [.71,.79]	.88	**.70**
Manipulative	Aggression	.72	.885	.859	.184 [.177-.191]	.954	.943	.117 [.110-.124]	.74 [.70,.78]	.77	.79
Perseverance	Dependability	.72	.918	.897	.241 [.233-.248]	.988	.984	.094 [.086-.102]	.75 [.73,.77]	.85	.82
Anxiety	Depression	.73	.948	.936	.214 [.207-.221]	.995	.993	.068 [.061-.076]	.76 [.74,.78]	.88	.90
Assertive	Leadership	.73	.851	.813	.174 [.167-.182]	.932	.913	.119 [.111-.127]	.72 [.68,.76]	**.69**	.72
Positivity	Gratitude	.73	.913	.891	.213 [.206-.221]	.969	.960	.129 [.121-.137]	.72 [.68,.76]	.84	.73
Sociability	Affability	.73	.890	.865	.223 [.216-.230]	.954	.943	.145 [.138-.152]	.73 [.69,.77]	.73	.85
Anxiety	Rumination	.73	.938	.925	.203 [.197-.210]	.988	.985	.091 [.084-.098]	.75 [.71,.79]	.88	.83
Manipulative	**Dishonest-Opportunism**	.74	.896	.870	.173 [.166-.181]	.962	.951	.106 [.098-.114]	.75 [.71,.79]	.77	**.70**
Procrastination	Distractibility	.74	.927	.909	.225 [.218-.233]	.984	.979	.107 [.099-.115]	.76 [.74,.78]	.83	.86
Altruism	**Empathy**	.76	.840	.800	.194 [.187-.202]	.929	.909	.131 [.124-.139]	.75 [.71,.79]	**.73**	**.73**
Dependability	**Honesty**	.76	.920	.902	.190 [.183-.196]	.964	.955	.128 [.121-.135]	.81 [.79,.83]	.82	**.78**
Anxiety	**Worry**	.76	.933	.918	.212 [.205-.218]	.980	.976	.115 [.109-.123]	.79 [.77,.81]	.88	**.76**
Aggression	**Rude**	.76	.907	.886	157 [.150-.164]	.952	.940	.114 [.107-.121]	.79 [.77,.81]	.79	**.70**
Social Confidence	Affability	.77	.897	.874	.245 [.238-.252]	.934	.918	.197 [.191-.204]	.82 [.80,.84]	**.79**	.85

Rumination	Sensitivity to Criticism	.78	.901	.879	.206 [.199-.213]	.961	.951	.130 [.123-.137]	.76 [.74,.78]	.83	.79
Depression	Inferiority	.78	.940	.926	.206 [.199-.212]	.980	.975	.120 [.113-.127]	.80 [.78,.82]	.90	.80
Positivity	**Surgency**	.79	.917	.899	.218 [.212-.225]	.954	.943	.165 [.158-.172]	.80 [.78,.82]	.84	**.78**
Worry	Rumination	.79	.927	.911	.185 [.178-.192]	.970	.962	.120 [.113-.127]	.80 [.78,.82]	**.76**	.83
Social Confidence	**Leadership**	.80	.895	.869	.208 [.212-.225]	.930	.911	.172 [.164-.179]	.78 [.76,.80]	.79	**.72**
Positvity	Depression	.80	.933	.918	.227 [.212-.225]	.947	.934	.203 [.196-.210]	.80 [.78,.82]	.84	.90
Aesthetics	**Sensitivity**	.81	.931	.912	.122 [.114-.131]	.952	.936	.104 [.095-.113]	.81 [.77,.85]	**.71**	**.61**
Affability	Social Boldness	.82	.925	.903	.234 [.225-.242]	.959	.946	.175 [.167-.184]	.81 [.79,.83]	.85	.82
Aggression	Provocative	.82	.938	.925	.160 [.153-.167]	.966	.957	.120 [.113-.127]	**.85[.83,.87]**	.79	.85
Provocative	**Rude**	**.85**	.950	.939	.133 [.127-.140]	.975	.969	.096 [.089-.103]	**.86[.84,.88]**	.85	**.69**
Social Confidence	**Social Boldness**	**.85**	.917	.893	.232 [.212-.225]	.946	.929	.189 [.180-.198]	**.83[.81,.85]**	.79	**.82**

Note: Scales typed in Boldface failed one or more of the statistical tests

General Discussion

The goal of the current research was to make meaningful strides towards the development of a comprehensive taxonomy of personality facets. We believe we have achieved this aim. Through an iterative, multi-stage, multi-method approach we have developed psychometrically robust, open access, scales that assess 70 essentially unidimensional, non-redundant, personality facets (Table S3.5). This Facet-level Multidimensional Assessment of Personality or Facet MAP, version 1, was developed, by analysing 1,772 personality items from many of the major omnibus personality inventories (e.g., NEO-PI-R; HEXACO, 16PF), by a conceptual review of 706 additional scales found within the literature, and a series of psychometric evaluations. The combination of multiple methods applied iteratively enabled us to balance the advantages and disadvantages of individual approaches. The Facet MAP builds upon and synthesises years of extant research concerning the assessment and structure of personality traits, addressing numerous calls for the systematic and empirical development of a more comprehensive list of personality facets (e.g., Condon, 2017; Condon et al., 2020; Goldberg, 1981; Hughes & Batey, 2017; Mõttus et al. 2020; Saucier et al., 2020).

The Facet MAP covers most existing facet scales within the literature and has approximately double the number of facets identified in most other taxonomies. Table 6 gives an illustration using the NEO-PI-R and SPI facet lists. This mapping was conducted at the level of scale definitions and items in order to minimize problems associated with Jingle-Jangle, especially because numerous facets from other inventories are multidimensional, covering content from more than one Facet MAP scale. Thus, we do not claim precise equivalence. Nevertheless, these mapping activities demonstrate how embedded the Facet MAP is within the extant literature and demonstrates the major strides it offers towards a comprehensive facet-level taxonomy.

Table 6

Equivalent facets from the Facet MAP, NEO-PI-R, and SPI estimated at the item-level to avoid

jingle-jangle issues with scale labels

NEO-PI-R	Facet MAP	SPI	Facet MAP
Neuroticism			
Anxiety	Blend: Anxiety & Worry	Anxiety	Blend: Anxiety & Worry
Hostility	Anger	Irritability	Anger
Depression	Depression	Well-being	Blend: Depression & Positivity
Self-consciousness	Blend: Sensitivity to criticism & Inferiority	Conformity	Blend: Need for Social Acceptance & Eccentricity
Impulsiveness	Blend: Self-control & Impetuous	Impulsivity	Impetuous
Vulnerability	Blend: Social Dependence & Self-efficacy	Emotional stability	No equivalent
		Emotional Expressivity	Personal Disclosure
Extraversion			
Warmth	Warmth		
Gregariousness	Sociability	Sociability	Sociability
Assertiveness	Blend: Assertive & Self-reliance		
Activity	Blend: Vigour & Procrastination	Easy-going	Vigour (-ve)
Excitement-Seeking	Blend on Sensation Seeking & Risk-aversion	Sensation-Seeking	Sensation Seeking
Positive Emotions	Positivity		
		Attention-seeking	Attention-seeking
		Humour	Humour
		Charisma	Social Boldness
Openness-to-Experience			
Fantasy	Blend: Fantasy & Imagination		
Aesthetics	Blend: Aesthetics & Absorption	Art Appreciation	Aesthetics
Feelings	Introspection	Introspection	Introspection
Actions	Novelty Seeking	Adaptability	Novelty Seeking
Ideas	Blend: Abstract Thinking and Intellectual Curiosity	Intellect	Blend: Intellectual Curiosity & Need for Cognition
Values	Blend: Traditionalism and Intolerance	Conservatism	Blend: Traditionalism & Spirituality
		Authoritarianism	Blend: Traditionalism & Punitive
		Creativity	Blend: Abstract Thinking, Imagination, Eccentricity & Fantasy
Agreeableness			
Trust	Suspicious	Trust	Suspicious
Straightforward	Blend: Honesty & Manipulative	Honesty	Blend: Honesty & Manipulative

Altruism	Altruism		
Compliance	Blend: Anger, Aggression, Forgiving, Stubborn	Irritability	Anger
Modesty	Grandiosity		
Tender-mindedness	Blend: Callous & Sympathy	Compassion	Empathy

Conscientiousness

Competence	Self-Efficacy		
Order	Blend: Orderliness & Planful	Order	Orderly
Dutifulness	Blend: Dependability & Dishonest-Opportunism	Industry	Procrastination (-ve)
Achievement-Striving	Achievement-striving		
Self-Discipline	Blend: Procrastination, Perseverance & Self-control	Self-control	Self-control
Deliberation	Blend: Impetuous & Planful		
		Perfectionism	Detail Conscious

Facets beyond the NEO-PI-R		**Facets beyond the SPI**	
Affability	Need for Cognition	Absorption	Gratitude
Attention-seeking	Need for Social Acceptance	Achievement Striving	Hedonism
Courageous	Personal Disclosure	Affability	Indecisive
Critical	Punitive	Aggression	Inferiority
Detail Conscious	Rumination	Altruism	Intolerance
Distractibility	Social Boldness	Assertive	Perseverance
Eccentricity	Spirituality	Callous	Planful
Emotion-based decision making	Tolerance for Ambiguity	Courageous	Positivity
Empathy	Vengeful	Critical	Risk-Aversion
Envy		Dependability	Rumination
Extrospection		Dishonest-Opportunism	Self-Efficacy
Fair		Distractibility	Self-Reliance
Gratitude		Emotion-Based Decision Making	Sensitivity to Criticism
Hedonism		Envy	Social Dependence
Humour		Extrospection	Stubborn
Indecisive		Fair	Tolerance for Ambiguity
Insecure Attachment		Fantasy	Vengeful
Intellectual Curiosity		Forgiving	Warmth
Intolerance		Grandiosity	

Because the Facet MAP provides the most comprehensive set of unique and unidimensional facets to-date, it should provide a useful reference for future work on personality assessment. For example, any ostensibly novel personality facet developed from this point could include explicit tests of its relationships with all of the conceptually relevant scales within the Facet MAP (see Hughes,

2018; Lawson & Robins, 2021). Anything truly novel can be added to the MAP and anything else can be discarded, reducing construct proliferation (Schwaba et al., 2020). Further, it might be the case that ostensibly novel personality scales/constructs are distinct from any single facet. However, a collection of facets might well explain the variance in such a scale. In such instances, we can begin to distinguish between 'elemental facets' and 'compound constructs'. For example, the 'dark triad' variables are 'compound constructs' that combine numerous facets contained within the Facet MAP (e.g., Lynam et al., 2011). Decomposing compound constructs into their constituent facets would again prevent construct proliferation, jingle-jangle, and allow for a more coherent science of human personality.

We also see numerous other uses of the Facet MAP. In some regard, most existing personality research could be meaningfully re-estimated with a larger facet list. Notable avenues in this regard include investigations of structural models of personality, cross-cultural patterns, developmental trajectories, and explanatory or predictive equations. In all cases, the more comprehensive facet-level descriptions are likely to lead to novel insights and, based on trends in existing research, greater levels of explanation than we typically see with current models (e.g., Hughes & Batey, 2017; Judge et al., 2013). Over time, using the open-access Facet MAP would enable the field to create redundancy-free databases of cumulative knowledge concerning facet-outcome correlations, functionally equivalent to the domain-level correlation databases, largely facilitated by the success of the 'Big Few' models (Condon et al., 2020). If the databases were large enough, the standard errors associated with such prediction equations could be very small and aid in the identification of the key facet predictors of outcomes. Conceivably such an approach could facilitate a more systematic accumulation of knowledge in psychological science.

Limitations and future research

All methodologies, statistical models, and conceptual decisions have inherent limitations. Studies relying predominantly on just one methodological approach are highly susceptible to such limitations. Thus, we adopted an iterative, multi-stage, multi-method approach to reduce the effects of any particular methodology. In many cases, this meant combining a range of statistical parameters (e.g., factor loadings, model fit etc.) with conceptual judgements (e.g., definitional review, semantic item review). We see the use of conceptual judgements in combination with rigorous statistical techniques as an overall strength of this work (Lawson & Robins, 2021). Taken together, evidence from one of the largest item-level analyses to-date, the systematic review of the personality assessment literature, and the rigorous empirical investigations of the newly developed scales, represent a meaningful step towards a comprehensive taxonomy of personality facets. Nevertheless, the Facet MAP remains preliminary and much further work is required, including cross-validation using samples that are even more diverse, the use of multiple-methods (to counteract single-method biases), and repeated assessments to assess re-test reliability, amongst all of the other useful validation practices (e.g., predictive validity, incremental predictive validity etc., Hughes, 2018). Regarding the work included within this paper, there are also several specific limitations worth considering.

First, it is likely that the Facet MAP misses some important facets. For example, due to the semantic grouping in Study 1, which was necessary to aid optimal factor recovery, facets might not have been identified during EFA because items were unable to coalesce when placed into different groups. However, the systematic search conducted in Study 2 went a long way to addressing this limitation. Specifically, when reviewing 706 facets from the literature we found surprisingly few additional facets beyond the base list developed in Study 1. Thus, although claims of comprehensiveness are premature, the facet MAP clearly captures a significant proportion of the extant literature. Nevertheless, we have specific plans to develop the Facet MAP, in at least four

ways. One, adopting a 'crowdsourcing' methodology. For example, circulating e-mails via psychological societies asking for missing scales. Two, conducting additional systematic searches for novel personality scales produced since Study 3 data collection was completed. Three, it would be useful to examine whether personality nuances (typically operationalised as single items or residualised item variance; Mõttus et al., 2017; 2020; Stewart et al., 2022) can indicate potentially omitted content. McCrae's (2015) original conception of nuances was that Five-Factor Model facets might be partitioned into smaller units (e.g., 'bitterness' and 'touchiness' might be different nuances of the 'angry hostility' facet, Mõttus et al., 2017). We agree and interpret this argument as evidence for the multi-dimensionality of many facets within existing inventories. Thus, nuances and the unidimensional facets identified here might well assess similar phenomena, with the two approaches being complementary in mapping narrow personality traits. Four, the SAPA project is now investigating between 6,000 and 13,000 personality items. Some of these items will likely be equivalent to those within the Facet MAP and some will form novel scales beyond. Thus, it should be possible to cross validate the Facet MAP and assess the remaining items to identify anything missing from version 1 of the Facet MAP.

Second, we might have duplicated facets within the Facet MAP. The extensive quantitative and qualitative checks for redundancy in Study 1 and Study 3 have gone some way to addressing this potential limitation. However, we have specific plans to assess this further. Most notably in the form of criterion associations which help identify the degree to which each facet explains unique variance in important social outcomes.

Finally, although the quantitative portions of Study 1 and Study 3 used larger and more diverse participant samples than are typical in studies of personality structure and scale development (e.g., DeYoung et al., 2007; Rouco et al., 2022; Saucier et al., 2020), they remain modest for the

46

analysis of such large item-sets. Thus, it is imperative that larger and more representative samples are collected.

Summary

The three studies presented here describe a programme of research designed to make meaningful strides towards the development of a comprehensive taxonomy of personality facets. Combining large-scale quantitative models and conceptual evaluations we were able to circumvent many of the issues concerning scale multidimensionality, construct proliferation and jingle-jangle. The resultant Facet MAP provides definitions and open-access, psychometrically robust scales to assess 70 narrow, unidimensional, and non-redundant personality facets. The Facet MAP contains scales equivalent to almost all scales present in major personality inventories, and in most cases, many more as well. We believe that version 1 of the Facet MAP will prove useful in reducing construct proliferation and confusion across the field, as well as facilitating important research regarding personality trait structure, developmental and cross-cultural trends, and explanatory models of life outcomes. Developing taxonomies of personality traits at any level of the hierarchy represents a huge undertaking; the Facet MAP was no different and would not have been possible without decades of previous high-quality research. As the Facet MAP continues to develop, we hope it will eventually represent a comprehensive taxonomy of personality facets.

Open Science Statement.

All unique data and code is available here:

https://osf.io/6mkuh/?view_only=c052f6c8168447f1bf8bf0c32689fb51

References

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied, 15*(2), 163.

Altemeyer, B. (1981). *Right-Wing Authoritariamsm.* University of Manitoba Press.

Anglim, J., & Grant, S. L. (2014). Incremental criterion prediction of personality facets over factors: Obtaining unbiased estimates and confidence intervals. *Journal of Research in Personality, 53*, 148-157.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Lawrence Erlbaum Associates, Inc.

Ashton, M. C., Jackson, D. N., Paunonen, S. V., Helmes, E., & Rothstein, M. G. (1995). The criterion validity of broad factor scales versus specific facet scales. *Journal of Research in Personality, 29*(4), 432-442.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150-166.

Ashton, M. C., Lee, K., & Son, C. (2000). Honesty as the sixth factor of personality: Correlations with Machiavellianism, primary psychopathy, and social adroitness. *European Journal of Personality, 14*(4), 359-368.

Booth, T., & Hughes, D. J. (2014). Exploratory structural equation modeling of personality data. *Assessment, 21*(3), 260–271.

Booth, T., & Murray, A. L. (2018). How factor analysis has shaped personality trait psychology. In *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development, Vols. 1-2* (pp. 933-951). Wiley Blackwell. https://doi.org/10.1002/9781118489772.ch29

British Psychological Society. (2018). *BPS Code of Ethics and Conduct*. https://cms.bps.org.uk/sites/default/files/2022-06/BPS%20Code%20of%20Ethics%20and%20Conduct.pdf

Brown, T. (2006a). CFA with equality constraints, multiple groups, and mean structures. *Confirmatory factor analysis for applied research, 1*(1), 236-319.

Brown, T. A. (2006b). Confirmatory factor analysis for applied research. New York, NY: The Guilford Press.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81.

Cattell, R. B., & Tsujioka, B. (1964). The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales1. *Educational and Psychological Measurement, 24*(1), 3-30.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319. https://doi.org/10.1037/1040-3590.7.3.309

Condon, D. M. (2017). The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model. https://doi.org/10.31234/osf.io/sc4p9

Condon, D. M., Roney, E., & Revelle, W. (2017). A SAPA project update: On the structure of phrased self-report personality items. *Journal of Open Psychology Data, 5*(1).

Condon, D. M., Wood, D., Mõttus, R., Booth, T., Costantini, G., Greiff, S., Johnson, W., Lukaszewski, A., Murray,A., Revelle, W., Aidan G. C. Wright, A. G. C., Matthias Ziegler, M., & Zimmermann, J. (2020). Bottom up Construction of a Personality Taxonomy. *European Journal of Psychological Assessment, 36*(6), 923-934. https://doi.org/10.1027/1015-5759/a000626

Conn, S. R., & Rieke, M. L. (Eds.). (1994). *The 16PF Fifth Edition Technical Manual.* Institute for Personality and Ability Testing.

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in webor internet-based surveys: Unnumbered graphic rating scales versus likert-type scales. *Educational and Psychological Measurement, 61*(4), 697-706.

Costa, P. T., & McCrae, R. R. (1992). *Manual for the Revised NEO Persoanoity Inventory (NEO-PIR) and NEO Five-Factor Inventory*. Odessa, FL: Psychological Assessment Resources

Costa, P. T., & McCrae, R. R. (2008). *The Revised Neo Personality Inventory (NEO-PI-R)*. Sage Publications, Inc.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(1), 7.

Couper, M. P. (2008). *Designing effective Web surveys*. Cambridge University Press.

Cox III, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of marketing research, 17*(4), 407-422.

De Raad, B., & Barelds, D. P. (2008). A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. *Journal of Personality and Social Psychology, 94*(2), 347.

de Winter, J., Dodou, D., & Wieringa, P. (2009). Exploratory Factor Analysis With Small Sample Sizes. *Multivariate Behavioral Research, 44*, 147-181. https://doi.org/10.1080/00273170902794206

de Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*(4), 695-710. https://doi.org/10.1080/02664763.2011.610445

DeVellis, R. F. (2016). *Scale Development: Theory and Applications* (4th ed.). Sage.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*(6), 1138.

DeYoung, C. G. (2015, 2015/06/01/). Cybernetic Big Five Theory. *Journal of Research in Personality, 56*, 33-58. https://doi.org/https://doi.org/10.1016/j.jrp.2014.07.004

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

Finstad, K. (2010). Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of usability studies, 5*(3), 104-110.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*(6), 1011-1027. https://doi.org/10.1037/0022-3514.80.6.1011

Flora, D. B., & Curran, P. J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods, 9*(4), 466-491. https://doi.org/10.1037/1082-989X.9.4.466

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research, 18*(1), 39-50.

Fox, J., & Dusa, A. (2022). *polycor: Polychoric and Polyserial Correlations*. https://CRAN.R-project.org/package=polycor

Goldberg LR (1981). Language and individual differences: The search for universals in personality lexicons In Wheeler L. (Ed.), Review of personality and social psychology (Vol. 2, pp. 141–165). Sage.

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

Goldberg, L. R. (2008). *The Eugene-Springfield community sample: Information available from the research participants* (ORI Technical Report, Issue.

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.

Goodman, J. K., Gabriele, & Paolacci, G. (2017). Crowdsourcing consumer research. *Journal of Consumer Research, 44*(1), 196-210.

Gough, H. G., & Bradley, P. (Eds.). (2002). *CPI™ manual* (3rd edition ed.). The Myers-Briggs Company.

Grucza, R. A., & Goldberg, L. R. (2007). The comparative validity of 11 modern personality inventories: Predictions of behavioral acts, informant reports, and clinical indicators. *Journal of Personality Assessment, 89*(2), 167-187.

Hattie, J. A. (1984). Decision criteria for assessing unidimensionality : An empirical study. *Multivariate Behavioral Research, 19*, 49-78. Hattie, J. A. (1985).

Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and ltenls. *Applied psychological measurement, 9*(2), 139-164.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104-121.

Hofstee, W. K., De Raad, B., & Goldberg, L. R. (1992). Integration of the big five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63*(1), 146.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.

Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Hughes, D. J., & Batey, M. (2017). Using personality questionnaires for selection. In H. Goldstein, E. Pulakos, J. Passmore, & C. Semedo (Eds.), The wiley blackwell handbook of the psychology of recruitment, selection and retention. Chichester: Wiley-Blackwell

Hughes, D. J. (2018). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, 751-779.

Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in human behavior, 28*(2), 561-569.

Irwing, P., & Hughes, D. J. (2018). Test Development. In *The Wiley Handbook of Psychometric Testing*. John Wiley & Sons Ltd. https://doi.org/10.1002/9781118489772.ch1

Jackson, D. N. (1994). *JPI-R: Jackson Personality Inventory*. Sigma Assesment Systems, Incorporated.

Jackson, D. N., Ashton, M. C., & Tomes, J. L. (1996). The six-factor model of personality: Facets from the Big Five. *Personality and Individual Differences, 21*(3), 391-402.

Jackson, D. N., Paunonen, S. V., Fraboni, M., & Goffin, R. D. (1996). A five-factor versus six-factor model of personality structure. *Personality and Individual Differences, 20*(1), 33-45.

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98*(6), 875.

Kelley, K., & Lai, K. (2018). Sample size planning for confirmatory factor models: Power and accuracy for effects of interest. In *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development, Vols. 1-2* (pp. 933-951). Wiley Blackwell. https://doi.org/10.1002/9781118489772.ch5

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*(3), 486-507.

Kline, P. (Ed.). (1993). *The handbook of psychological testing*. Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling, 3rd ed*. Guilford Press.

Lawson, K. M., & Robins, R. W. (2021). Sibling constructs: What are they, why do they matter, and how should you handle them?. Personality and Social Psychology Review, 25(4), 344-366

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*(2), 329-358.

Lee, K., Ashton, M. C., Hong, S., & Park, K. B. (2000). Psychometric properties of the nonverbal personality questionnaire in Korea. *Educational and Psychological Measurement, 60*(1), 131-141.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Little, R. J., & Rubin, D. B. (2002). Bayes and multiple imputation. *Statistical analysis with missing data*, 200-220.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading MA: Addison-Wesley

Lounsbury, J. W., Sundstrom, E., Loveland, J. M., & Gibson, L. W. (2003). Intelligence,"Big Five" personality traits, and work drive as predictors of course grade. *Personality and Individual Differences, 35*(6), 1231-1239.

Lynam, D. R., Gaughan, E. T., Miller, J. D., Miller, D. J., Mullins- Sweatt, S., & Widiger, T. A. (2011). Assessing the basic traits associated with psychopathy: Development and validation of the Elemental Psychopathy Assessment. Psychological Assessment, 23, 108–124. https://doi.org/10.1037/a0021146

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36*(4), 611-637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84.

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

McCrae, R. R. (2015). A more nuanced view of reliability: Specificity in the trait hierarchy. *Personality and Social Psychology Review, 19*(2), 97-112.

McNeish, D., An, J., & Hancock, G. R. (2018, 2018/01/02). The Thorny Relation Between Measurement Quality and Fit Index Cutoffs in Latent Variable Models. *Journal of Personality Assessment, 100*(1), 43-52. https://doi.org/10.1080/00223891.2017.1281286

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology, 112*(3), 474.

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., & Lukaszewski, A. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits. *European Journal of Personality, 34*(6), 1175-1201.

Mulaik, S. A. (2018). Fundamentals of common factor analysis. *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, 209-251.

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.* http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (Eighth ed.). Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*(6), 609-626.

Oxford English Dictionary (2015). Oxford University Press.

Paunonen, S. (2002). Design and construction of the Supernumerary Personality Inventory. *Research Bulletin, 763*.

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*(3), 524-539. https://doi.org/10.1037/0022-3514.81.3.524

Paunonen, S. V., Haddock, G., Forsterling, F., & Keinonen, M. (2003). Broad versus narrow personality measures and the prediction of behaviour across cultures. *European Journal of Personality, 17*(6), 413-433. https://doi.org/10.1002/per.496

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the big five? Plenty! *Journal of Personality, 68*(5), 821-835.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*(4), 741.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? *Sociological methodology*, 73-104.

Revelle, W. (2016). How to: Use the psych package for factor analysis and data reduction. *Evanston, IL: Northwestern University, Department of Psychology*.

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. https://CRAN.R-project.org/package=psych

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment, 31*(12), 1395.

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research, 14*(4), 403-414.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145-154.

Rönkkö, M., & Cho, E. (2020, 2022/01/01). An Updated Guideline for Assessing Discriminant Validity. *Organizational Research Methods, 25*(1), 6-14. https://doi.org/10.1177/1094428120968614

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.

Rouco, V., Cengia, A., Roberts, R., Kemper, C., & Ziegler, M. (2022). The Berlin Multi-Facet Personality Inventory: An IPIP-based measure of Big Five personality facets. *Psychological Test Adaptation and Development*.

Salgado, J. F., Moscoso, S., Sanchez, J. I., Alonso, P., Choragwicka, B., & Berges, A. (2015). Validity of the five-factor model and their facets: The impact of performance measure and facet residualization on the bandwidth-fidelity dilemma. *European Journal of Work and Organizational Psychology, 24*, 325-349

Saucier, G., & Iurino, K. (2020). High-dimensionality personality structure in the natural language: Further analyses of classic sets of English-language trait-adjectives. *Journal of Personality and Social Psychology, 119*(5), 1188-1219. https://doi.org/10.1037/pspp0000273

Saucier, G., Iurino, K., & Thalmayer, A. G. (2020). Comparing predictive validity in a community sample: High–dimensionality and traditional domain–and–facet structures of personality variation. *European Journal of Personality*, *34*(6), 1120-1137

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*(2), 147.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23-74.

Schwaba, T., Rhemtulla, M., Hopwood, C. J., & Bleidorn, W. (2020). A facet atlas: Visualizing networks that describe the blends, cores, and peripheries of personality structure. *PLOS ONE, 15*(7), e0236893.

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods, 19*(1), 80-110.

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.

Soutter, A. R. B., & Mõttus, R. (2021). Big five facets' associations with pro-environmental attitudes and behaviors. *Journal of Personality, 89*(2), 203-215.

Stewart, R. D., Mottus, R., Seeboth, A., Soto, C. J., & Johnson, W. (2022). The finer details: The predictability of life outcomes from Big Five domains, facets and nuances. Journal of Research in Personality, 90, 167-182. DOI: 10.1111/jopy.12660

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589-617.

Stout, W. (2002). Psychometrics: From practice to theory and back.. *Psychometrika, 67*, 485-518.

Tellegen, A. (1991). Personality traits: Issues of definition, evidence, and assessment. In Thinking clearly about psychology: Essays in honor of Paul E. Meehl, Vol. 1: Matters of public interest; Vol. 2: Personality and psychopathology (pp. 10–35). Minneapolis: University of Minnesota Press

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. *The SAGE handbook of personality theory and assessment, 2*, 261-292.

VandenBos, G. R. (2007). *APA dictionary of psychology*. American Psychological Association.

van Buuren, S., & Groothuis-Oudshoorn, K. (2021). *Multivariate Imputation by Chained Equations (MICE)*. https://github.com/amices/mice

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.

Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement and evaluation in counseling and development, 34*(3), 166-176.

Willis, G. B., Schechter, S., & Whitaker, K. (2000). *A comparison of cognitive interviewing, expert review and behavior coding: What do they tell us?* American Statistical Association.

Wood, D., Nye, C. D., & Saucier, G. (2010). Identification and measurement of a more comprehensive set of person-descriptive trait markers from the English lexicon. Journal of Research in Personality, 44(2), 258-272.

Ziegler, M., Booth, T., & Bensch, D. (2013). Getting entangled in the nomological net. Thoughts on validity and conceptual overlap. European Journal of Psychological Assessment, 29, 157–161. http://dx.doi.org/10.1027/1015-5759/a000173.

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfalls and loopholes. *European Journal of Psychological Assessment, 31*, 231-237.