NovoPsych

"NovoPsych helps mental health services use psychometric science to improve client outcomes"

Development of a Short Personality Assessment: The NovoPsych Five Factor Personality Scale - 30-item version.

Ben Buchanan[1 2] and David Hegarty[2]

2023

[1] Monash University, Faculty of Medicine, School of Psychological Sciences

[2] NovoPsych Pty Ltd, Australia

Corresponding author email: david@novopsych.com

Abstract

Objective: Personality assessments provide valuable insights into diverse life outcomes and stability over time, aiding in understanding individuals' behaviours and traits. However, traditional lengthy assessments can deter participation and lead to response biases; a more concise version offers a strategic balance between comprehensive insight and improved respondent engagement.
Method: This research created a shortened personality assessment, the NovoPsych Five Factor Personality Assessment - 30-item version (NFFPS-30), from the International Personality Item Pool - Neuroticism, Extraversion, Openness - 120 item version (IPIP-NEO-120) using 410,195 responses from a publically available international sample. Both second-order and bi-factor models were tested using Confirmatory Factor Analysis to determine model fit. The best fitting models were selected and the highest loading items in each facet were then used to create the NFFPS-30.
Results: Confirmatory Factor Analysis confirmed that the bi-factor models fit the data best and the resultant facets and factors of the NFFPS-30 were highly correlated with the facets and factors of the IPIP-NEO-120. Percentiles were created to allow for the NFFPS-30 to be used for clinical or research purposes within an Australian context.
Conclusions: It was concluded that the NFFPS-30 could be a valuable tool in the assessment of personality for clients within a clinical practice or research setting.

Key Words: *Individual differences, inventory, personality, personality assessment.*

Buchanan, B., & Hegarty, D., (2023). Development of a Short Personality Assessment: The NovoPsych Five Factor Personality Scale - 30-item version.

Published 11 September 2023

Personality factors hold significance for a range of life outcomes, with demonstrated predictive validity in areas like relationships and well-being (Roberts et al., 2007) and correlate with objective outcomes like income and education (Kajonius & Carlander, 2017). Personality factors are gaining prominence due to the individualistic context of modern society (Skirbekk & Blekesaune, 2014) and they exhibit stability and predictable development across the lifespan (Briley & Tucker-Drob, 2014).

The prevalent framework for personality measurement is the five-factor model (FFM; McCrae, 2010), encompassing Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N). Recurring correlations among these factors have reignited interest in FFM's structure (Strus et al., 2014), especially given its use in DSM-5 (American Psychiatric Association, 2013) and integrating common and abnormal factors (Markon et al., 2005). Empirical support bolsters FFM's utility in the conceptualisation of psychopathology including personality disorders (Miller, 2012).

Building upon the FFM framework and drawing upon the International Personality Item Pool (Goldberg et al., 2006), the International Personality Item Pool - Neuroticism, Extraversion, Openness - 120 item version (IPIP-NEO-120) offers a comprehensive representation of personality factors within a concise item set (Johnson, 2014). To enhance the precision of personality measurement, the IPIP-NEO-120 includes facets - subtraits within each domain which exemplifies behaviours, affects, and cognitions within each factor (Zillig et al., 2002). In essence, the IPIP-NEO-120 serves as an exemplar of a well-crafted personality assessment that resonates with both theoretical rigour and practical versatility, rendering it a valuable tool for understanding and evaluating individual differences.

However, the length of the IPIP-NEO-120 could be a barrier for clients. Reducing the length of the IPIP-NEO-120 holds several compelling rationales grounded in both practical and theoretical considerations. Firstly, the demand for brevity aligns with the inherent time constraints that clinicians and clients face. Lengthy assessments may deter participation or lead to hasty responses, potentially compromising the reliability and accuracy of gathered data. Secondly, the pursuit of conciseness addresses cognitive fatigue, wherein extended engagement with a complex task can diminish the quality of responses and lead to response biases (Robins et al., 2001). By shortening the assessment's length, the cognitive burden on clients is mitigated, fostering sustained attention and ensuring more thoughtful and considered responses.

Personality factors encompass multifaceted nuances, and it is imperative for assessment tools to accurately capture this complexity while also remaining manageable for respondents. Lengthy assessments can result in redundancy, wherein similar items measure the same construct repeatedly. In contrast, a compact assessment encourages the selection of items that maximise distinctiveness and relevance, enhancing the discriminative power of the instrument. Additionally, the strategic curation of items in a shortened version can facilitate targeted assessment of specific personality facets, factors, or constructs of particular interest, yielding focused insights that align with the assessment's intended

purpose. This study aimed to shorten the IPIP-NEO-120 while balancing between obtaining comprehensive information and ensuring respondent engagement, accurate measurement, theoretical soundness and the practical utility of the instrument. This article outlines our goal of shortening the IPIP-NEO-120 to a 30-item version using structural CFA models, resulting in the NovoPsych Five Factor Personality Scale - 30-item version (NFFPS-30).

Method

Sample. We utilised a substantial publicly available international sample (N = 619,150), who had completed the IPIP-NEO-120. The dataset is readily available in a publicly accessible psychology research repository (see Johnson, 2020). Initial data from the repository was first filtered for some data errors where responses were zero for some questions (given questions responses need to be one to five) and if any rows contained a zero in a response, the whole row was removed (resultant n = 410,376). Data that exhibited a pattern of non-unique values across all values (i.e., selecting the same response across the whole test) was also removed (resultant n = 410,195). Finally, the age of clients was then used to remove data for clients who were below the age of 16 (resultant n = 385,735). The final sample consisted of 41% male (average age 25.8) and 59% female (average age 25.5) with an age range from 16 to 95 years old. As reported by Johnson (2014), data were acquired through an online survey platform hosted on a website, which provided participants with feedback on their FFM personality following their completion of all items. This sample was entirely voluntary, from diverse demographic backgrounds, and was drawn from various sources including word-of-mouth, search engines, and informal networks.

Instrument. The IPIP-NEO-120 is a publicly available FFM assessment (Johnson, 2014), using 120 items from the IPIP (Goldberg et al., 2006). Built on open-source items correlated with NEO-PI-R (Costa & McCrae, 1995), the five factor scales comprised four correlated items per facet, with six facets per factor. Individual items were rated on a 1 (almost never) – 5 (almost always) scale, summarised into 30 facet traits (Min = 4, Max = 20) and five factors (Min = 24, Max = 120). Overall, mean facet scale reliability was $\alpha = .78$; only four facets (13%) were below $\alpha = .70$ (Johnson, 2014).

Statistical Method. Data analysis was performed in RStudio (Version 2022.07.2+576; RStudio Team, 2022) using R (Version 4.2.0; R Core Team, 2022). The analyses were based on the six facet traits within each factor in the IPIP-NEO-120 and the main objective was achieved by testing hierarchical structural CFA models to determine which item had the highest loading per facet. The models were based on established FFM literature (Costa & McCrae, 1995), and tested the loadings between manifest items, latent trait facets, and respective latent factor domain. As per Kajonius & Johnson (2019), two models were tested for each of the latent factor domains (see Figure 1). Firstly, a second-order model, which is the original and most used hierarchical structure (Costa & McCrae, 1995), with the general latent factor domain at the top, loaded by six facet traits, which in turn were loaded by 24 item measurement items (Figure 1a). Secondly, a bi-factor model, which didn't allow covariances between

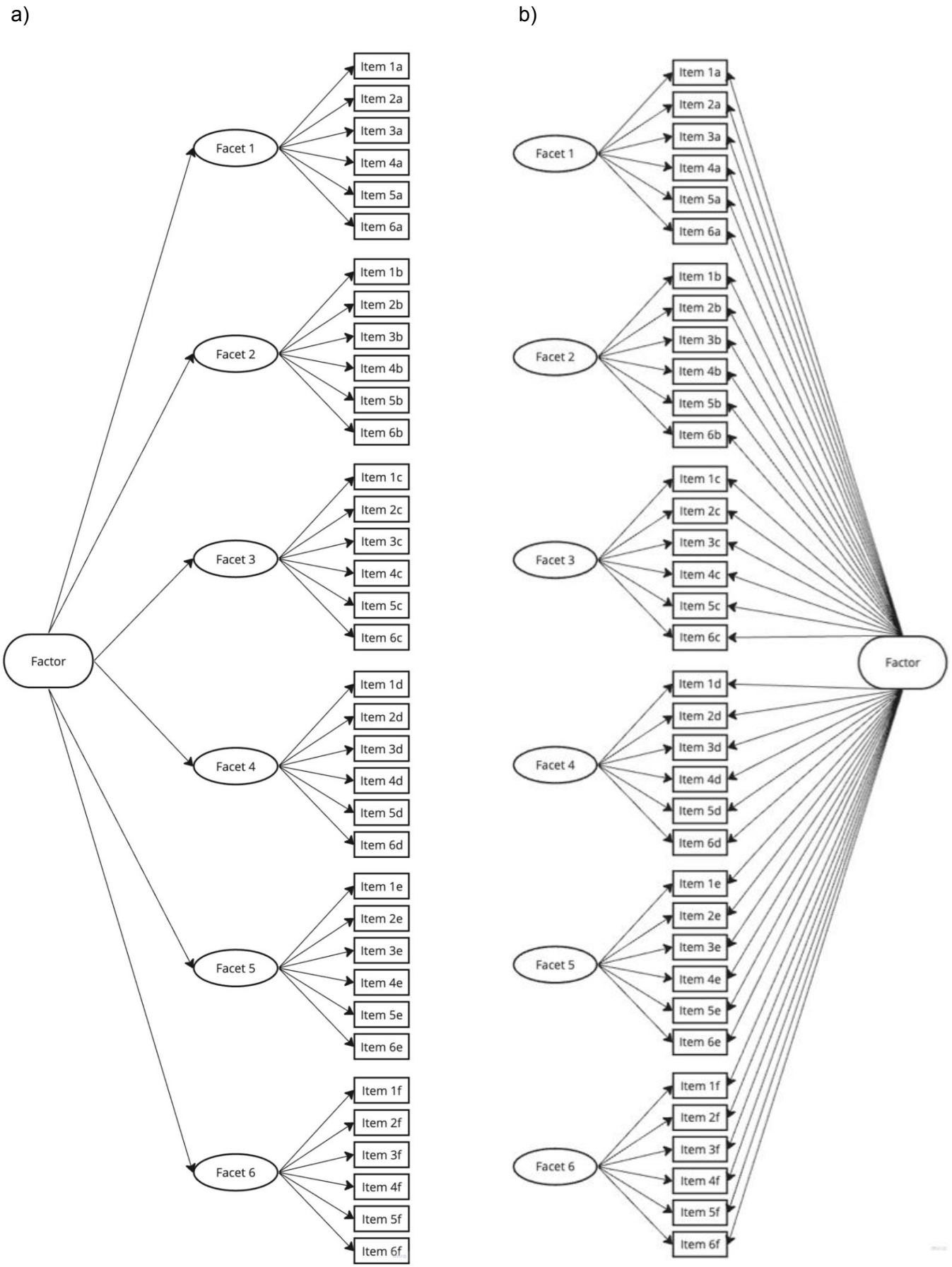

Figure 1. Path diagrams for the tested models for each of the five factors within the IPIP-NEO-120. a) The structure of the FFM second-order models. b) The structure of the FFM bi-factor models.
Note. FFM = Five Factor Model.

factors, where both the general latent factor and specific latent facets are loaded onto directly by measurement items (Figure 1b). These CFA models were conducted in a SEM-framework, extracted with Maximum Likelihood, using the lavaan package (Rosseel, 2012). The models were initially attempted as unconstrained, without covarying error terms, but were adjusted as necessary (as long as it still made conceptual sense) to improve model fit.

Results

Confirmatory Factor Analysis. All five CFA factor structure models, one for each domain, were conducted in two ways: Second-order models and bi-factor models. Initially, all parameters were unconstrained, and conducted without modifications (e.g., covariances between errors). Only two of the second-order models provided an overall satisfactory model fit: Conscientiousness and Neuroticism (see Table 1 for fit statistics). However, for these two second-order models there were negative latent variable variances for with the Self-Discipline and Anxiety facets, respectfully, being the problems. Overall, the bi-factor models provided a better fit than the second-order models. However, for the Conscientiousness, Extraversion, and Neuroticism bi-factor models, it was necessary to constrain variances of factors to one and thereby freely estimate the loading of the first item in order to get the model to fit. In addition, the Neuroticism factor required further modifications to enable model fit, with one item from the Vulnerability facet and all items from the Anxiety facet having to load on the neuroticism factor alone.

Table 1.
Second-Order and Bi-Factor Model Fit Statistics

Factor Domain	Second-Order Model (Fig. 1a)	Bi-Factor Model (Fig. 1b)
Openness	$\chi^2(246) = 265{,}978$; RMSEA = 0.05; TLI = 0.87; CFI = 0.89; AIC = 26,708,856	$\chi^2(228) = 183{,}421$; RMSEA = 0.05; TLI = 0.91; CFI = 0.92; AIC = 26,626,334
Conscientiousness	$\chi^2(246)^{\dagger} = 265{,}054$; RMSEA = 0.05; TLI = 0.93; CFI = 0.93; AIC = 23,244,688	$\chi^2(228)^{\ddagger} = 216{,}174$; RMSEA = 0.05; TLI = 0.93; CFI = 0.95; AIC = 23,195,843
Extraversion	$\chi^2(246) = 494{,}136$; RMSEA = 0.07; TLI = 0.86; CFI = 0.88; AIC = 24,631,399	$\chi^2(228)^{\ddagger} = 354{,}504$; RMSEA = 0.06; TLI = 0.89; CFI = 0.91; AIC = 24,491,803
Agreeableness	$\chi^2(246) = 409{,}525$; RMSEA = 0.07; TLI = 0.86; CFI = 0.88; AIC = 24,338,659	$\chi^2(228) = 237{,}933$; RMSEA = 0.05; TLI = 0.91; CFI = 0.93; AIC = 24,167,104
Neuroticism	$\chi^2(246)^{\dagger} = 360{,}472$; RMSEA = 0.06; TLI = 0.90; CFI = 0.91; AIC = 26,024,137	$\chi^2(233)^{\S} = 334{,}764$; RMSEA = 0.06; TLI = 0.90; CFI = 0.91; AIC = 25,998,456

Note. RMSEA = Root Mean Square Error of Approximation; TLI = Tucker Lewis Index ; CFI = Comparative Fit Index; AIC = Akaike Information Criteria. Fit indices (χ^2, RMSEA, TLI, CFI, AIC) are reported based on Figure 1a and Figure 1b, without improvement modifications (e.g., covariations between errors or between facets). † some estimated latent variable variances were negative; ‡ had to constrain variances of factors to one and thereby freely estimate the loading of first item; § had to remove an item in the Vulnerability facet (Item 26) and all the Anxiety facet items (Items 1, 31, 61, 91) and have all of these load onto the Neuroticism factor alone.

Fit indices for both the second-order and bi-factor models could have been improved by allowing one or more selected error variances to covary within each trait facet. Furthermore, the negative latent variables that were found for Conscientiousness and Neuroticism were able to be solved by allowing one or more

selected error variances to covary within each facet. Both of these solutions also improved model fit slightly, however, it was decided to keep the models as simple as possible.

Loadings on facet levels for the bi-factor models were high overall. The average facet loading in the bi-factor models was 0.66, although there were a few weaker loadings (defined as $\beta < .50$), such as facet traits Self Discipline and Self Efficacy in the Conscientiousness factor, Friendliness in the Extraversion factor, and Vulnerability in the Neuroticism factor. Notably, in each of these cases the item loaded more strongly onto the general factor trait instead. The loadings on factor levels were overall lower than those on the facet level, with the average item loading 0.43. However, given the bi-factor nature of these models, it was clear that all items loaded onto both the facet and factor, just at varying levels for each item. That is, when an item had a weaker loading at the facet level, it was higher at the factor level and vice versa.

Table 2.

Factor Correlations Between the NFFPS-30 Factors and IPIP-NEO-120 Factors.

Factor	*t* †	Correlation (95% CI)
Openness	1065.8	.86 (.86 - .86)***
Conscientiousness	1575.1	.93 (.93 - .93)***
Extraversion	1308.2	.90 (.90 - .90)***
Agreeableness	1186.3	.89 (.89 - .89)***
Neuroticism	1569.4	.93 (.93 - .93)***

Note. NFFPS-30 = NovoPsych Five Factor Personality Scale - 30-item version; IPIP-NEO-120 = International Personality Item Pool - Neuroticism, Extraversion, Openness - 120 item version; CI = Confidence Interval; † Degrees of freedom on all correlations was 385,733; *** $p < .001$

It is notable that the Neuroticism bi-factor model required one item from the Vulnerability facet and all items from the Anxiety facet to load at the factor level only for the model to fit. Each of these items did appear to be somewhat similar in appearing to measure some aspect of anxiety or worrying in general: Item 1 (Anxiety facet): *Worry about things*; Item 26 (Vulnerability facet): *Panic Easily*; Item 31 (Anxiety facet): *Fear for the worst*; Item 61 (Anxiety facet): *Am afraid of many things*; and Item 91 (Anxiety facet): *Get stressed out easily*. It was unclear exactly why none of the items from the Anxiety facet appeared to fit the bi-factor model and were simply loaded onto the general Neuroticism factor. However, it was apparent that in the initial bi-factor model Item 1 in the Anxiety facet had a factor loading that was considerably higher than all other items (15.5 compared to 0.007, 0.002, and 0.006 respectively) and it had a very large (and negative) variance. Attempting to constrain this factor loading or its starting value did not help. Even when this item was removed from the Anxiety facet the model wouldn't fit due to negative loadings for other items within the facet, which was not theoretically consistent. As a result of

Table 3.

Facet Items with their Facet and Factor Loadings and Correlations with IPIP-NEO-120 Facets.

Facet (Factor)	IPIP-NEO-120 Item Number	Facet Loading	Factor Loading	NFFPS-30 Item Number	Correlation (95% CI)[†]
Adventureness (O)	78	0.69	0.26	17	.78 (.78 - .78)
Artistic Interests (O)	98	0.62	0.53	25	.81 (.81 - .81)
Emotionality (O)	103	0.68	0.23	26	.76 (.75 - .76)
Imagination (O)	63	0.84	0.28	13	.82 (.82 - .82)
Intellect (O)	113	0.65	0.48	29	.80 (.80 - .80)
Liberalism (O)	88	0.80	0.25	22	.80 (.80 - .80)
Achievement Striving (C)	80	0.51	0.61	18	.85 (.85 - .86)
Cautiousness (C)	30	0.73	0.40	8	.87 (.87 - .87)
Dutifulness (C)	105	0.68	0.46	27	.77 (.77 - .77)
Orderliness (C)	70	0.73	0.45	15	.87 (.87 - .87)
Self Discipline (C)	85	0.39	0.66	21	.80 (.80 - .81)
Self Efficacy (C)	95	0.41	0.58	24	.78 (.78 - .78)
Activity Level (E)	17	0.79	0.25	6	.81 (.81 - .81)
Assertiveness (E)	12	0.77	0.39	4	.87 (.87 - .87)
Cheerfulness (E)	117	0.64	0.49	30	.83 (.82 - .83)
Excitement Seeking (E)	82	0.70	0.10	19	.76 (.76 - .76)
Friendliness (E)	62	0.48	0.71	12	.82 (.82 - .82)
Gregariousness (E)	7	0.72	0.57	3	.82 (.82 - .83)
Altruism (A)	49	0.60	0.34	10	.79 (.79 - .79)
Cooperation (A)	44	0.62	0.57	9	.73 (.73 - .74)
Modesty (A)	84	0.91	0.11	20	.84 (.84 - .84)
Morality (A)	69	0.61	0.59	14	.83 (.83 - .83)
Sympathy (A)	29	0.70	0.38	7	.79 (.79 - .79)
Trust (A)	4	0.76	0.30	1	.86 (.86 - .86)
Anger (N)	6	0.75	0.44	2	.88 (.88 - .88)
Depression (N)	71	0.65	0.59	16	.87 (.87 - .87)
Immoderation (N)	111	0.60	0.31	28	.75 (.75 - .75)
Anxiety (N)	91	N/A[‡]	0.79	23	.81 (.81 - .81)
Self Consciousness (N)	16	0.66	0.39	5	.79 (.79 - .79)
Vulnerability (N)	56	0.39	0.65	11	.77 (.77 - .77)

Note. O = Openness; C = Conscientiousness; E = Extraversion; A = Agreeableness; N = Neuroticism; [†] All correlations were statistically significant at $p < .001$ and CI = Confidence Interval; [‡] The Anxiety facet item loads onto the Neuroticism factor alone.

this issue, it was decided to keep the highest loading item on the Neuroticism factor that was formerly in the Anxiety facet (Item 91) and keep it as a measure of Anxiety to ensure consistency with the IPIP-NEO-120.

Construct Validity. As a result of the factor analysis, the 30 items that loaded onto each facet (including the Anxiety item that loaded onto the Neuroticism factor) were included to create the NFFPS-30. The resultant 30-items had a good balance of positively and negatively scored items with an even split of 15 items each. To determine the construct validity of the NFFPS-30, correlations were calculated between the new facet and factor-level items and the IPIP-NEO-120 facets and factors. As seen in Table 2, the correlation between NFFPS-30 factors and IPIP-NEO-120 factors were high to very strong, with an average correlation of .90. As presented in Table 3, all correlations between NFFPS-30 facets and IPIP-NEO-120 facets were high, with an average correlation of .81 and all correlations greater than .73. All of the aforementioned correlations were statistically significant ($p < .001$; df = 385,733).

Norming / Percentiles. Using the facets and factors for the NFFPS-30 identified by the CFA process, percentiles for each facet and factor were calculated using the cNORM package (Version 3.0.2; Lenhard & Lenhard, 2021). This method of norming estimates percentiles on the basis of the raw data without requiring assumptions about the distribution of the raw data. This method minimises bias arising from sampling and measurement error, while handling marked deviations from normality, addressing bottom or ceiling effects and capturing almost all of the variance in the original norm data sample (Lenhard & Lenhard, 2021). It is beyond the scope of this article to present all percentile tables in an exhaustive fashion, but Table 4 presents the percentiles for the five factors. Given the factor scores were based upon the six facets that made up each factor (i.e., scores between 6 and 30), these scores were consistently well distributed and non-problematic (see Table 5).

Given the evidence for minor cross-cultural differences in personality (Kajonius & Mac Giolla, 2017; Rentfrow & Gosling, 2013; Rentfrow et al., 2015), it was necessary to create norms that might be of clinical or research use locally. Therefore, to calculate percentiles for the NFFPS-30 for use in an Australian context, the sample was limited to those where the respondent was in Australia (resultant n = 14,163). These respondents were made up of 5,252 males (between the age of 16 - 95) and 8,911 females (between the age of 16 - 88). Given that the respondent age was skewed positively, with a mean age of 26.9, the data was categorised into age groups to allow approximately equal sized groups (n ~ 2,000) for comparison. The resultant age groups were 16-17 year old (n = 2,509), 18-19 year olds (n = 2,279), 20-21 year olds (n = 1,624), 22-25 year olds (n = 1,742), 26-30 year olds (n = 2,032), 31-39 year olds (n = 2,128), and 40 year olds plus (n = 1,849). Given that for all facets and factors on the Australian sample, there were differences in the distribution of scores by age and by gender (all p's < .001), it was decided to break up the percentiles for clinical or research use. However, when dealing with very large sample sizes, the probability of achieving statistical significance tends to be elevated. Consequently,

Table 4.

Percentiles for the Five Factors of the NFFPS-30.

Factor Score	Percentiles				
	Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**
6					0.9
7					1.6
8		0.1	0.1		2.6
9		0.2	0.2	0.1	4.2
10	0.1	0.4	0.4	0.2	6.5
11	0.3	0.8	0.7	0.4	9.7
12	0.6	1.4	1.5	0.7	14.0
13	1.2	2.5	2.8	1.2	19.3
14	2.3	4.3	4.9	2.1	25.8
15	4.2	6.8	8.2	3.5	33.2
16	7.3	10.4	12.9	5.7	41.3
17	11.8	15.2	19.2	9.0	49.8
18	18.0	21.3	27.1	13.8	58.3
19	25.9	28.7	36.4	20.2	66.4
20	35.3	37.0	46.5	28.5	73.9
21	45.7	46.0	56.9	38.6	80.4
22	56.4	55.3	66.8	50.1	85.8
23	66.7	64.2	75.7	62.2	90.1
24	75.9	72.4	83.1	73.8	93.3
25	83.4	79.6	88.8	83.7	95.7
26	89.3	85.5	93.0	91.3	97.3
27	93.5	90.2	95.9	96.1	98.4
28	96.2	93.6	97.7	98.7	99.1
29	98.0	96.0	98.8	99.7	99.5
30	99.0	97.6	99.4	99.9	99.7

many researchers consider effect size calculations to provide a more informative perspective in such cases (Lin, Lucas, & Shmueli, 2013). The differences between factor scores by gender and by age are presented in Tables 6 and 7 respectively. The percentiles for the NFFPS-30 were calculated for each facet and factor based upon the aforementioned age groupings and gender (male, female, and combined). Despite the facet percentiles being calculated based upon only one question with five potential responses, the distribution of responses were not overly skewed and percentiles have a high degree of utility for interpretation purposes. Only the Dutifulness and Achievement Striving facets demonstrate somewhat of a skewed distribution (positively skewed) for all genders but only in the 40 years of age and older age group. These percentiles for all facets and factors are available upon request from the corresponding author.

Table 5.

Distribution of Scores on the Five Factors of the NFFPS-30

Scale	M	SD	Skew	Kurtosis	Internal Reliability α
Openness	21.4	3.7	-0.22	-0.17	0.43
Conscientiousness	21.4	4.3	-0.25	-0.37	0.69
Extraversion	20.3	3.8	-0.34	-0.05	0.54
Agreeableness	21.7	3.3	-0.54	0.54	0.46
Neuroticism	17.0	4.6	0.10	-0.48	0.70

As demonstrated in Table 6, there were significant differences between males and females on each of the five factors measured by the NFFPS-30 in the Australian data. Despite all comparisons being statistically significant ($p < .01$), some of these differences are obviously more meaningful than others. For example, it could be argued that the difference between males and females on both the Conscientiousness and Extraversion factors was negligible (i.e., $d < 0.2$), whereas the differences between males and females on both the Agreeableness and Neuroticism factors were more substantial, even if they might still be classified as a ‘small’ effect size. In all of the factors outlined in Table 6, females scored higher than males.

Table 6.
Mean Differences in Personality Traits across Gender for the NFFPS-30 Australian Data

Factor	Gender[†]	Mean	SD	Interpretation
Openness	Male	21.48	3.58	A significant difference between males and females on the Openness factor with females scoring higher than males (*d* = 0.26; 95% CI [0.23, 0.30])
	Female	22.40	3.42	
Conscientiousness	Male	20.63	4.42	A significant difference between males and females on the Conscientiousness factor with females scoring higher than males (*d* = 0.05; 95% CI [0.01, 0.08])
	Female	20.84	4.33	
Extraversion	Male	20.43	3.97	A significant difference between males and females on the Extraversion factor with females scoring higher than males (*d* = 0.09; 95% CI [0.06, 0.13])
	Female	20.79	3.79	
Agreeableness	Male	21.24	3.53	A significant difference between males and females on the Agreeableness factor with females scoring higher than males (*d* = 0.41; 95% CI [0.37, 0.44])
	Female	22.58	3.12	
Neuroticism	Male	16.0	4.68	A significant difference between males and females on the Neuroticism factor with females scoring higher than males (*d* = 0.40; 95% CI [0.36, 0.43])
	Female	17.83	4.60	

[†] Males (n = 5,245), females (n = 8,911).

As demonstrated in Table 7, there were significant differences between age groups for most of the five factors of the NFFPS-30 for the Australian data. The exceptions were within the Extraversion factor where there were only three significant differences and between the 16-17 age group and 18-19 age group on the Openness factor where there was no significant difference. Of note, on both the Extraversion and Openness factors, even the significant differences could be considered negligible in size (i.e., $d < 0.2$). The biggest difference between the age groups was on the Conscientiousness factor between the 16-17 age group and the 40 plus age group, where there was a large effect size (i.e., $d > 0.8$), with Conscientiousness increasing significantly with age. There was also a medium effect size (i.e., $d > 0.5$) for the change between 16-17 and the 40 plus age group on Neuroticism, with Neuroticism decreasing significantly with age.

Table 7.

Mean Differences in Personality Traits across Age Groups for the NFFPS-30 Australian Data

Factor	Age Group[†]	Mean	SD	Interpretation
Openness	16 -17	21.73	3.40	A significant difference between the 16-17 age group and all other age groups apart from the 18-19 age group. Minimum *d* = 0.12; 95% CI [0.06, 0.18] (16-17 *v* 20-21 age group). Maximum *d* = 0.16; 95% CI [0.1, 0.23] (16-17 *v* 22-25 age group).
	18-19	21.86	3.44	
	20-21	22.14	3.43	
	22-25	22.30	3.56	
	26-30	22.23	3.51	
	31-39	22.23	3.48	
	40+	22.06	3.70	
Conscientiousness	16 -17	18.91	4.15	A significant difference between the 16-17 age group and all other age groups. Minimum *d* = 0.16; 95% CI [0.10, 0.22] (16-17 *v* 18-19 age group). Maximum *d* = 1.09; 95% CI [1.02, 1.15] (16-17 *v* 40+ age group).
	18-19	19.56	4.03	
	20-21	19.92	4.16	
	22-25	20.70	4.32	
	26-30	21.42	4.19	
	31-39	22.13	4.15	
	40+	23.29	3.87	
Extraversion	16 -17	20.88	3.92	A significant difference between the 16-17 age group and 18-19 age group (*d* = -0.08; 95% CI [-0.02, -0.13]), between the 16-17 age group and the 31-39 age group (*d* = -0.11; 95% CI [-0.05, -0.16]), and between the 16-17 age group and the 40+ age group (*d* = -0.19; 95% CI [-0.13, -0.25])
	18-19	20.59	3.79	
	20-21	20.91	3.95	
	22-25	20.77	3.93	
	26-30	20.8	3.83	
	31-39	20.47	3.81	
	40+	20.14	3.75	
Agreeableness	16 -17	21.54	3.64	A significant difference between the 16-17 age group and all other age groups. Minimum *d* = 0.11; 95% CI [0.05, 0.18] (16-17 *v* 20-21 age group). Maximum *d* = 0.38; 95% CI [0.32, 0.44] (16-17 *v* 40+ age group).
	18-19	22.05	3.35	
	20-21	21.94	3.37	
	22-25	22.02	3.23	
	26-30	22.12	3.2	
	31-39	22.25	3.21	
	40+	22.85	3.09	

Table 7 (continued).

Neuroticism	16 -17	18.19	4.38	A significant difference between the 16-17 age group and all other age groups. Minimum *d* = -0.08; 95% CI [-0.02, -0.13] (16-17 *v* 18-19 age group). Maximum *d* = -0.58; 95% CI [-0.52, -0.64] (16-17 *v* 40+ age group).
	18-19	17.85	4.39	
	20-21	17.70	4.68	
	22-25	17.20	4.65	
	26-30	16.67	4.83	
	31-39	16.58	4.79	
	40+	15.56	4.80	

† 16-17 age group (n = 2,509), 18-19 age group (n = 2,279), 20-21 age group (n = 1,624), 22-25 age group (n = 1,742), 26-30 age group (n = 2,032), 31-39 age group (n = 2,128), 40 plus age group (n = 1,849).

Discussion

This study aimed to shorten the IPIP-NEO-120 into a 30-item assessment using CFA as a method to determine the items that loaded highest onto each facet. As a result of the CFAs performed, the bi-factor models generally provided a good fit to the data, apart from an issue with model fit for the Neuroticism factor. As a result of these fit issues, the Anxiety facet should be interpreted with caution and considered as a more general Neuroticism measure. It is not fully understood why there was no Anxiety facet in the bi-factor model but it would appear that in this sample the Anxiety items were merely functioning as a measure of an individual's typical degree of distress as opposed to anxiety specifically, and therefore, the Anxiety items were better associated with Neuroticism in general - with all types of emotional distress.

As a result of the bi-factor model facet loadings, the NFFPS-30 was created. Due to model fit, the structure of the NFFPS-30 is theoretically sound in that it measures personality traits at the facet level and then all items within the facet are loaded onto a more generalised latent factor. This 30-item measure (see Table 3 for IPIP-NEO-120 items that are included in the NFFPS-30) appears to validly measure the same facets and factors that are measured by the IPIP-NEO-120, due to the high correlations between the facets and factors. Clearly it would be beneficial to have additional measures of concurrent validity, and this could be a focus in future research, but the NFFPS-30 does appear to have good content, construct, and face validity.

Percentiles were calculated for the NFFPS-30 for use within an Australian context and these percentiles are split up into age groups and by gender. The percentiles provide added utility for using the NFFPS-30 within a clinical or research context as they allow for a normative comparison between the respondent’s scores and respondents of a similar age and gender. Normative comparisons offer valuable insights into variations among individuals in terms of personality traits and abilities. Notably, these comparisons can serve as a valuable tool for monitoring advancements within the expanding array of programs rooted in personality change. These programs are designed to facilitate deliberate personality transformation

through the establishment of precise behavioural objectives and tasks. (e.g., Hudson et al., 2019; Hudson & Roberts, 2014; Martin et al., 2014). Furthermore, the norms developed for the NFFPS-30 demonstrate that Conscientiousness increases with age and Neuroticism decreases with age, a finding that is both consistent with the growing consensus that personality traits continuously change across the lifespan (Damian et al., 2019; Twomey & Johnson, 2022; Wagner, Ram, Smith, & Gerstorf, 2016) and adds to the external validity of the NFFPS-30. The norms for the NFFPS-30 also found some small differences in personality factors by gender, with the largest differences in Agreeableness and Neuroticism. Again this is a finding that is consistent with extant research (e.g., see review by Lippa, 2010) and adds to the validity and potential clinical and research usefulness of the NFFPS-30.

Limitations. One limitation of this shortened personality assessment is with the measure of facets by using only one item. Having multiple items to measure a facet of personality in an assessment is useful for enhancing measurement reliability, validity, and accuracy. Multiple items allows for each item to provide a distinct perspective on the construct, reducing the impact of random response fluctuations or measurement errors. This redundancy also allows for capturing different aspects of the construct, reducing the risk of item-specific biases and increasing the overall robustness of the assessment. However, given that lengthy assessments may also diminish the quality of responses and lead to response bias (Robins et al., 2001), there is obviously a trade-off that is occurring when deciding upon a lengthy personality assessment, such as the IPIP-NEO-120 or a shortened version, like the NFFPS-30. Nevertheless, single-item measures have been shown to be reliable and valid previously (e.g., Robins et al., 2001) and given the high correlations between the NFFPS-30 facets and factors and the IPIP-NEO-120 facets and factors, it is believed that the NFFPS-30 can provide a reliable, valid, and accurate measure of personality traits.

As outlined by Kajonius & Johnson (2019), the extensive sample used for this study is unlikely to be a representative nationwide sample. Due to the active volunteering of respondents, a reasonable assumption is that several of the facet personality traits, such as in Emotionality, Intellect, and Altruism, were likely considerably higher than normal, since these traits are known to characterise people interested in psychology (Vedel, 2016). Another limitation with the sample is the age of respondents - it does not capture a good representation of older adults and thus might miss some significant personality trends that are common amongst older adults (e.g., see Wagner et al., 2016).

It is also noted that the Cronbach's alpha results for each of the five factors in the NFFPS-30 were relatively low (see Table 5). This is not unsurprising given that alpha increases as a function of test length (Schmitt, 1995) and given that as part of the CFA process, emphasis was placed upon ensuring that the item from each facet that loaded most heavily upon that facet was kept (regardless of its loading upon the factor). An alternative approach to shorten the IPIP-NEO-120 that could have kept high factor internal reliability, would have been to simply select the items that loaded most heavily at the factor level, but this would have potentially removed valuable facet level personality measures. Given that the

NFFPS-30 is derived from a bi-factor model and there are two sources of variance (at the facet and factor level), it could be problematic to even use Cronbach's alpha as a measure of internal reliability (Ziegler & Bäckström, 2016). Nevertheless, the internal reliability of the NFFPS-30 factors are lower than those of the IPIP-NEO-120, which were 0.82, 0.91, 0.89, 0.86, and 0.90 for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism respectively. However, in relation to low internal reliability for scales, Schmitt (1995, p. 351 - 352) states that, "[w]hen a measure has other desirable properties, such as meaningful content coverage of some domain and reasonable unidimensionality, this low reliability may not be a major impediment to its use". Therefore, even though the internal reliability of the factor items for the NFFPS-30 are reasonably low, we believe given its high correlations with IPIP-NEO-120 at both the factor and facet level, that this does not detract from the usefulness of the NFFPS-30 as a quick and reliable tool for measuring personality.

Conclusion. The development of the NovoPsych Five Factor Personality Scale - 30-item version (NFFPS-30) through rigorous CFA modelling from the IPIP-NEO-120 holds substantial utility for both clinical and research domains. This shortened yet psychometrically sound measure efficiently captures core personality traits, facilitating efficient assessment while maintaining robustness, enabling practitioners and researchers to gain valuable insights into individuals' personality profiles without the burden of lengthy assessments.

References

American Psychiatric Association [APA]. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC, USA: Author.

Briley, D. A., & Tucker-Drob, E. M. (2014). Genetic and environmental continuity in personality development: A metaanalysis. Psychological Bulletin, 140(5), 1303-1331. https://doi.org/10.1037/a0037091

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. Journal of Personality Assessment, 64(1), 21-50. https://doi.org/10.1207/s15327752jpa6401_2

Damian, R. I., Spengler, M., Sutu, A., & Roberts, B. W. (2019). Sixteen going on sixty-six: A longitudinal study of personality stability and change across 50 years. Journal of Personality and Social Psychology, 117(3), 674–695. https://doi.org/10.1037/ pspp0000210

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. https://doi.org/10.1016/j.jrp.2005.08.007

Hudson, N. W., Briley, D. A., Chopik, W. J., & Derringer, J. (2019). You have to follow through: Attaining behavioural change goals predicts volitional personality change. Journal of Personality and Social Psychology, 117(4), 839–857. https://doi.org/ 10.1037/pspp0000221

Hudson, N. W., & Roberts, B. W. (2014). Goals to change personality traits: Concurrent links between personality traits, daily behavior, and goals to change oneself. Journal of Research in Personality, 53, 68–83. https://doi.org/10.1016/j.jrp.2014.08.008

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the IPIP-NEO-120. Journal of Research in Personality, 51, 78-89. https://doi.org/10.1016/j.jrp.2014.05.003

Johnson, J. A. (2020). Johnson's IPIP-NEO data repository. Accessed at: https://osf.io/tbmh5/

Kajonius, P. J., & Johnson, J. A. (2019). Assessing the Structure of the Five Factor Model of Personality (IPIP-NEO-120) in the Public Domain. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment*, *15*(2), 260–275. https://doi.org/10.5964/ejop.v15i2.1671

Kajonius, P., & Mac Giolla, E. (2017). Personality traits across countries: Support for similarities rather than differences. PLoS One, 12(6), e0179646. https://doi.org/10.1371/journal.pone.0179646

Lenhard, W., & Lenhard, A. (2021). Improvement of Norm Score Quality via Regression-Based Continuous Norming. Educational and Psychological Measurement, 81(2), 229–261. https://doi.org/10.1177/0013164420928457

Lin, M., Lucas, H. C., & Shmueli, G. (2013). Research commentary – Too big to fail: Large samples and the p-value problem. Information Systems Research, 24(4), 906–917. https://doi.org/10.1287/isre.2013.0480

Lippa, R.A. (2010), Gender Differences in Personality and Interests: When, Where, and Why?. Social and Personality Psychology Compass, 4: 1098-1110. https://doi.org/10.1111/j.1751-9004.2010.00320.x

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. Journal of Personality and Social Psychology, 88(1), 139-157. https://doi.org/10.1037/0022-3514.88.1.139

Martin, L. S., Oades, L. G., & Caputi, P. (2014). Intentional personality change coaching: A randomised controlled trial of participant selected personality facet change using the FiveFactor Model of Personality. International Coaching Psychology Review, 9(2), 196–209.

Miller, J. D. (2012). Five-factor model personality disorder prototypes: A review of their development, validity, and comparison to alternative approaches. Journal of Personality, 80(6), 1565-1591. https://doi.org/10.1111/j1467-1591.

McCrae, R. R. (2010). The place of the FFM in personality psychology. Psychological Inquiry, 21(1), 57-64. https://doi.org/10.1080/10478401003648773

Rentfrow, P. J., & Gosling, S. D. (2013). Divided we stand: Three psychological regions of the United States and their political, economic, social, and health correlates. Journal of Personality and Social Psychology, 105(6), 996–1012. https://doi.org/10.1037/a0034434

Rentfrow, P. J., Jokela, M., & Lamb, M. E. (2015). Regional personality differences in Great Britain. PLoS One, 10(3), e0122245. https://doi.org/10.1371/journal.pone.0122245

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. Personality and Social Psychology Bulletin, 27(2), 151–161. https://doi.org/10.1177/0146167201272002

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. Perspectives on Psychological Science, 2(4), 313-345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Schmitt, N. (1996). Uses and abuses of coefficient alpha. Psychological Assessment, 8(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Skirbekk, V., & Blekesaune, M. (2014). Personality traits increasingly important for male fertility: Evidence from Norway. European Journal of Personality, 28(6), 521-529. https://doi.org/10.1002/per.1936

Strus, W., Cieciuch, J., & Rowiński, T. (2014). The circumplex of personality metatraits: A synthesizing model of personality based on the big five. Journal of Personality and Social Psychology, 18(4), 273-286. https://doi.org/10.1037/gpr0000017

Twomey, C., & Johnson, J. A. (2022). Open-source personality trait norms for the United Kingdom and Ireland. European Journal of Psychological Assessment, 38(6), 432–439. https://doi.org/10.1027/1015-5759/a000644

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

Wagner, J., Ram, N., Smith, J., & Gerstorf, D. (2016). Personality trait development at the end of life: Antecedents and correlates of mean-level trajectories. Journal of personality and social psychology, 111(3), 411–429. https://doi.org/10.1037/pspp0000071

Ziegler, M., & Bäckström, M. (2016). 50 facets of a trait—50 ways to mess up?[Editorial]. European Journal of Psychological Assessment, 32(2), 105–110. https://doi.org/10.1027/1015-5759/a000372

Zillig, L. M. P., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in big 5 personality inventories. Personality and Social Psychology Bulletin, 28(6), 847-858. https://doi.org/10.1177/0146167202289013