Original Research Article

Mapping Chinese Personality: An Assessment of the Psychometric Properties of the NEO-PI-3 in Monolingual and Bilingual Studies

Assessment
2023, Vol. 30(7) 2031–2049

DOI: 10.1177/10731911221126921
journals.sagepub.com/home/asm
SAGE

Michelle Yik[1], Irene N. L. Sze[1], Felity H. C. Kwok[1], and Shiang-yi Lin[1]

Abstract

The NEO-PI-3 is a fourth-generation instrument that has been shown to improve the psychometrics and readability of its immediate precedent, the NEO-PI-R. We examined the psychometric properties of the Chinese versions of the NEO-PI-R and NEO-PI-3 using three datasets (*Ns* = 913, 299, 403) collected using both monolingual and bilingual designs. The Chinese NEO-PI-3 scales displayed a five-factor structure in which the facets had the highest loadings on their intended factors. The structure demonstrated strong invariance across both languages (English vs. Chinese) and gender groups, maintained high test-retest reliability, and attained slightly better internal consistency than the NEO-PI-R. We also examined the affective underpinnings of personality factors and well-being measures using the Chinese Circumplex Model of Affect. Consistent with past findings, Neuroticism and Extraversion were most related to affect, while Satisfaction with Life and Subjective Happiness shared the affective core of pleasant feelings and medium arousal. Based on these results, the Chinese NEO-PI-3 appears to be a sound instrument to measure personality in Chinese communities and to compare personality across cultures.

Keywords

NEO-PI-3, NEO-PI-R, Chinese Circumplex Model of Affect, bilingual studies, Chinese personality, measurement invariance

> The five-factor model has provided a unified framework for trait research; it is the Christmas tree on which findings of stability, heritability, consensual validation, cross-cultural invariance, and predictive utility are hung like ornaments.
>
> (Costa & McCrae, 1993, p. 302)

Much of what psychologists mean by "personality" can be succinctly summarized by the five-factor model of personality (FFM; McCrae & John, 1992; see Goldberg, 1981). The five factors, also known as the "Big Five," are Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). In a narrow sense, the FFM represents an umbrella of replicable factor structures resulting from hundreds of validation studies that have been conducted in different cultures (McCrae & Costa, 2010) and using different measurement devices (McCrae & John, 1992; Soto & John, 2017). In a broader sense, psychologists are now moving beyond the descriptive structure of the FFM to the five-factor theory of personality (McCrae & Allik, 2002). This theory promises to catalyze an integrated understanding of personality, to organize a myriad of empirical findings into a coherent story, and to establish connections between personality and other human conditions (McCrae & Costa, 2010; McCrae, Terracciano, et al., 2005; Terracciano et al., 2005). Meta-analyses have shown robust relationships between the Big Five and physical and mental health (Strickhouser et al., 2017), life satisfaction (Anglim et al., 2020), and academic success (Poropat, 2009).

The NEO inventories are among the most popular measures of the FFM. They have been translated into more than 50 languages, and the five-factor structure has been replicated in many of these languages in both self-reports and observer reports. These inventories have a hierarchical structure, organizing traits systematically (five domains with six facets each). McCrae and Costa (2010) provided a bibliography of more than 2,500 articles that use these

[1]Division of Social Science, Hong Kong University of Science and Technology

Corresponding Author:
Michelle Yik, Division of Social Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.
Email: Michelle.Yik@ust.hk

inventories in personality, industrial/organizational, and clinical research. This level of coverage has persuaded researchers and practitioners that the NEO scales provide an instrument with strong psychometric properties and consistent interpretations across the globe.

In the present study, we examined the scales' utility in mapping personality in the Chinese community, which is the largest ethnic group in the globe and represents 18% of the planet's population (Chen et al., 2019). To this end, we conducted a psychometric evaluation of a Chinese translation of the latest version of the NEO scales. After English, Chinese is the most widely spoken language in the world (Eberhard et al., 2021).

NEO Research in Chinese Communities

To conduct cross-cultural research on personality, researchers need a common assessment tool. To this end, they often use an imposed etic approach (Berry, 1969), in which scales developed and validated in English are "imported" for use in another language. For instance, McCrae, Costa, and Yik (1996) translated the NEO scales from English into Chinese, administered the translated scales to Chinese participants, and made inferences about the personalities of Chinese people using the English concepts.

The NEO-PI-R (Costa & McCrae, 1992) was first imported for use in Chinese communities in the 1990s. In these studies, the five-factor structure was replicated in some samples but not in others, although test-retest reliability was consistently high in all five domains (McCrae, Costa, & Yik, 1996; McCrae et al., 1998; Wu et al., 2008; Yang, 2010; Yang et al., 1999). Internal consistency scores varied across studies and facets. Yang (2010) reported adequate internal consistency at both the domain and facet levels among mainland Chinese respondents, while Yang et al. (1999) reported low consistency ($<$ .50) for several O and A facet scales (see also Wang et al., 2005). Wu et al. (2008) reported high consistency at the domain level for all five domains but low consistency for some E, O, and A facets among their Taiwanese respondents. In a meta-analysis of data collected in mainland China, Luo and Dai (2011) found that the internal consistency of the Chinese version of NEO-FFI (Costa & McCrae, 1992) was lower than that of the original English version.

Researchers have also examined the structure of the facet scales and extended this research into different populations. Testing the N facet scales in both clinical and nonclinical samples in mainland China, Xi et al. (2018) found support for the original six-facet structure, with acceptable internal consistency and test-retest reliability. Using differential item functioning, Dai et al. (2010) assessed the measurement invariance of the C facet scales in a U.S. sample and a mainland Chinese sample. Evidence was found for measurement invariance across these scales, supporting their use in cross-cultural comparisons.

Using data from mainland China, Hong Kong, and Taiwan, Cheung et al. (2008) studied the factor structure underlying the NEO-FFI (Costa & McCrae, 1992) and the Revised Chinese Personality Assessment Inventory (CPAI-2). Five factors were mapped: the first four resembled N, O, E, and A, while the fifth consisted of several items defining Interpersonal Relatedness and Dependability in the CPAI-2 along with some NEO C items.

Studies using the NEO scales in Chinese populations have therefore yielded mixed results in replicating the intended five-factor structure and attaining acceptable internal consistency (see Laajaj et al., 2019). Improvements made to the latest version of the NEO scales (the NEO-PI-3; McCrae & Costa, 2010) have been shown to improve its psychometric properties in predominantly White, educated, industrialized, rich, democratic ("WEIRD") communities (Henrich et al., 2010). This study aimed to determine whether this improvement held in a non-WEIRD (Chinese) community.

Development of the NEO-PI-3

Over the last four decades, several revisions have been made to improve the reliability and validity of the NEO scales. The NEO Inventory (NEO-I; McCrae & Costa, 1983) was first developed to assess the N, E, and O domains, each of which had corresponding facet scales. The domains A and C were developed subsequently and included in the NEO Personality Inventory (NEO-PI; Costa & McCrae, 1985). The revised NEO-PI (NEO-PI-R; Costa & McCrae, 1992) added facet scales to the A and C domains.

The NEO-PI-R (Costa & McCrae, 1992) has been widely used for more than 30 years in a range of populations and cultures. To test the replicability of its structure in different communities, the American normative structure, which was obtained from 1,000 college-educated Euro-American participants, has typically been used for cross-validation purposes. Despite the widespread use of the NEO-PI-R in previous research, some items were consistently found to be problematic in adolescent and adult samples (see McCrae, Terracciano, & 79 Members of the Personality Profiles of Cultures Project, 2005, for a review). For example, it was commonly reported that the English version of the NEO-PI-R included items with difficult terms (e.g., "fastidious" in C2: Order and "panhandlers" in A6: Tender-Mindedness) that were not comprehensible to respondents with limited literacy. Low coefficient alphas were also found in certain facet scales, and individual items with low item-facet correlations were identified as problematic. In total, McCrae, Costa, and Martin (2005) identified 48 problematic items of which 37 were chosen for replacement. In 2010, McCrae and Costa introduced the fourth generation (G4) of the NEO scales (viz., NEO-PI-3; McCrae & Costa, 2010) with the

objective of making the items more accessible to a wider population. Eleven items were replaced in the C domain, eight in the A domain, and six in each of the remaining domains. Item changes occurred in 19 of the 30 facets, with the highest number of replacements introduced in A6: Tender-Mindedness and C1: Competence.

Compared with their NEO-PI-R counterpart items, the modified items were found to be easier to understand and attained better psychometric properties in a predominantly White, high-achieving American adolescent sample (McCrae, Costa, & Martin, 2005). Readability and item-facet correlations were improved, while factor structure, correlates, and cross-observer agreement were virtually equivalent to those of the NEO-PI-R. At the facet level, the revisions made the most notable improvements in A6: Tender-Mindedness and C2: Order. Based on these results, the NEO-PI-3 appears to be a more effective measure than the NEO-PI-R: not only does it retain high reliability and validity, but it can also be administered to a wider range of populations, including respondents as young as 12 years of age and those with a reading level as low as Grade 5 (McCrae & Costa, 2010).

Adaptations of the NEO-PI-3 to Other Languages

The improved properties of the original English NEO-PI-3 had yet to guarantee the same improvements in the translated versions. To test whether the G4 scales made similar improvements in other languages, De Fruyt et al. (2009) took the lead by using observer ratings of adolescents aged 12 to 17 to compare the psychometric characteristics of the NEO-PI-R and the NEO-PI-3 in 18 different languages, including Chinese. Participants were asked to complete both the NEO-PI-R and the 37 modified items in the NEO-PI-3. Overall, the replacements improved the item-total correlations across languages, with marked improvement in the English-speaking language group. The internal consistency of the 19 facets with modified items was slightly improved in both English-speaking and non-English-speaking samples. The mean differences between the two versions were small, suggesting that these two NEO inventories are comparable.

Translated versions of the self-report NEO-PI-3 were also tested, including versions in Czech (Hrebícková, 2008), Greek (Fountoulakis et al., 2014), Canadian French (Le Corff & Busque-Carrier, 2016), Estonian (De Vries et al., 2016), South African English (Quy, 2011), and Swedish (Källmen et al., 2016). Acceptable internal consistency was reported in most samples, although the observed structural validity was less encouraging in some (see Källmen et al., 2016; Quy, 2011). In the Estonian sample, De Vries et al. (2016) examined self-other agreement in personality and reported low levels of agreement (.12 to .54).

Adaptations of the NEO-PI-3 in different languages have therefore had mixed success, with the results suggesting that complexities are introduced by culture and language when adapting a measurement tool developed in English to another target language (Chen, Benet-Martínez, & Ng, 2014; McCrae et al., 1998). In the present study, we validated the NEO-PI-3 and NEO-PI-R using three datasets of Chinese participants in Hong Kong. The participants self-reported their personality using either the NEO-PI-R or the NEO-PI-3. Both monolingual and bilingual designs were used.

NEO Correlates

To map the relationships between personality and affect, Yik et al. (2002; see also Yik, 2010a; Yik & Russell, 2001, 2004) correlated current affect, tapped by a circumplex model, with the NEO-FFI (Costa & McCrae, 1992). This mapping was conducted for two large samples of Canadian students and for five samples of participants who spoke different languages (English, Spanish, Chinese, Japanese, and Korean). Overall, those high on N tended to experience unpleasant (neither activated nor deactivated) affect, while those high on E tended to experience pleasantly aroused affect (e.g., *excited*). Kuppens et al. (2017) tested the moderating effect of personality on the correlation between valence and arousal and found that N and E were strongly related to the valence-arousal relationship (see also Yik et al., 2022). In the present study, we used Yik's (2009) Chinese Circumplex Model of Affect (CCMA) scales to examine the affective core of the G4 scales.

The NEO scales have also been found to be good predictors of subjective well-being. Life satisfaction has been shown to be negatively related to N but positively related to E, A, and C (McCrae & Costa, 1991; Wood et al., 2009; Yik et al., 2011). When examining the relationships between the 30 facets and well-being measures, over 80% of the correlations in the N, E, and C domains were found to be significant, compared with less than 50% in the O and A domains (Siegler & Brummett, 2000). In the present study, we tested the pattern of correlations between the G4 scales, Diener et al.'s (1985) Satisfaction with Life Scale, and Lyubomirsky's (2008) Subjective Happiness Scale.

Overview of the Present Study

In this study, we examined the psychometric properties of the Chinese NEO-PI-3. Using exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009), we validated the five-factor structure of the Chinese NEO-PI-3. The Chinese G4 scales displayed the five-factor structure with reasonably good internal consistency and invariance across gender groups (Datasets 2 and 3). They also demonstrated strict invariance across languages

Table 1. Overview of the Three Datasets.

Dataset	*N*	Measure		Goal
		Time 1	Time 2	
1 (monolingual design)	913	Chinese NEO-PI-R	N/A	(a) To test the structural validity of the Chinese NEO-PI-R (b) To generate the Chinese NEO-PI-R normative structure
2 (bilingual design)	299	English (or Chinese) NEO-PI-3	Chinese (or English) NEO-PI-3	(a) To test the structural validity of the Chinese NEO-PI-3 (b) To test the equivalence of the English and the Chinese versions of the NEO-PI-3
3 (monolingual design)	403	Chinese questionnaires: NEO-PI-3	Chinese questionnaires: NEO-PI-3 CCMA scales SWLS SHS	(a) To test the structural validity of the Chinese NEO-PI-3 (b) To test the test-retest reliability of the Chinese NEO-PI-3 (c) To examine the affective underpinnings of the NEO-PI-3 and its relationship with well-being measures

Note. NEO-PI-R = NEO Personality Inventory-Revised; NEO-PI-3 = NEO Personality Inventory-3; CCMA = Chinese Circumplex Model of Affect; SWLS = Satisfaction with Life Scale; SHS = Subjective Happiness Scale.

(Dataset 2), providing the first piece of evidence to support the English-Chinese equivalence of the NEO-PI-3. We also examined the affective underpinnings of the NEO-PI-3 and their relationship with well-being measures (Dataset 3) and found that the valence dimension of the affect circumplex served well as an intersection between affect and personality. Table 1 gives an overview of the three datasets.

Method

Participants and Procedures

The participants in this study were undergraduate students at a university in Hong Kong. University admission in Hong Kong requires that matriculating students demonstrate proficiency in both English and Chinese in standard language examinations. All of the participants in the present study were therefore considered fluent in both languages. To create Dataset 1, 913 participants (583 women; M_{age} = 19.96, SD_{age} = 1.18) completed the online version of the Chinese NEO-PI-R at home. To create Datasets 2 and 3, each participant visited a computer laboratory twice and completed a battery of online surveys. In Dataset 2, 299 participants (146 women; M_{age} = 21.12, SD_{age} = 1.06) were randomly assigned to complete the English (or Chinese) version of the NEO-PI-3 at Time 1 and the Chinese (or English) version at Time 2. In Dataset 3, 403 participants (222 women; M_{age} = 20.28, SD_{age} = 1.36) completed the Chinese NEO-PI-3 in both sessions. They also completed affect and well-being measures at Time 2. Unless otherwise stated, the questionnaires were in traditional Chinese.[1]

NEO Scales in Datasets 1 to 3

In all of the datasets, we included a self-report measure (Form S) of either the NEO-PI-R or NEO-PI-3, both of which are 240-item questionnaires designed to measure the FFM (Costa & McCrae, 1992). The responses were provided on a 5-point rating scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

Other Measures in Dataset 3

In addition to completing the Chinese NEO-PI-3, the participants completed several other scales.

Chinese Circumplex Model of Affect (CCMA) Scales. The CCMA scales (Yik, 2009) comprise 48 items designed to measure 12 affect segments, 1 o'clock (o'c) through 12 o'c, each of which is measured using four items. As shown in Figure 1, these scales are built along the horizontal axis of valence (0° vs. 180°) and the vertical axis of arousal (90° vs. 270°), each with different proportions of valence and arousal. The participants were asked to describe their feelings in the current moment on a 5-point rating scale ranging from 1 (*not at all*) to 5 (*extremely*). The score for each segment was the average of its four constituent items. Cronbach's alpha for each segment ranged from .75 (6 o'c) to .91 (5 o'c), which is comparable with the reliability estimates reported in previous studies (Yik, 2009, 2010a)

Satisfaction with Life Scale (SWLS). The SWLS (Diener et al., 1985) consists of five statements designed to measure life satisfaction. Our participants indicated the extent of

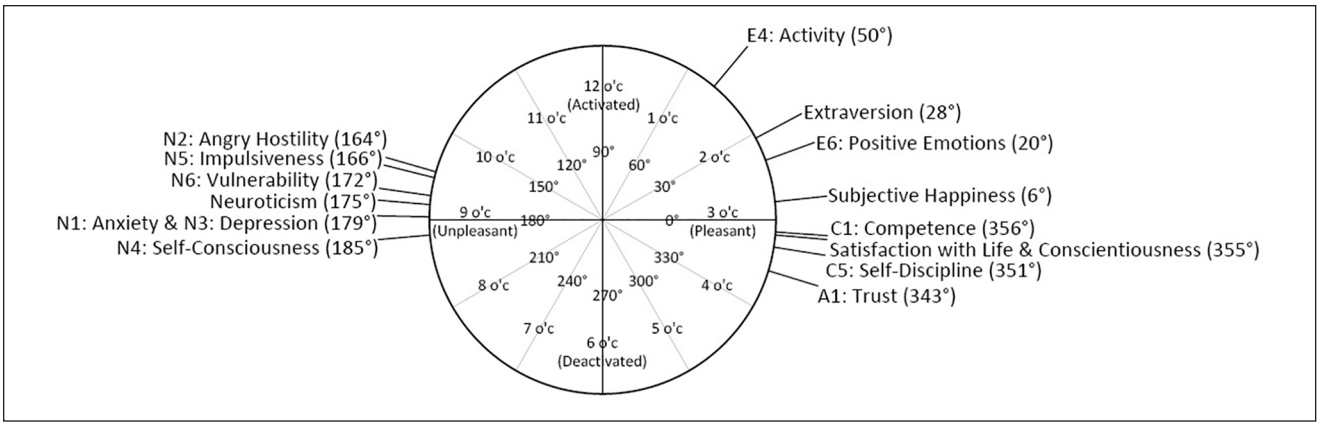

Figure 1. Placing 10 NEO-PI-3 Domain/Facet Scales and 2 Well-Being Scales Within the CCMA With the Cosine Wave Method ($N = 403$).
Note. NEO-PI-3 = NEO Personality Inventory-3; CCMA = Chinese Circumplex Model of Affect.

their agreement with each statement on a 7-point scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Their SWLS score was the mean of the five statements. Cronbach's alpha for this scale was .88.

Subjective Happiness Scale (SHS). The SHS (Lyubomirsky, 2008) consists of four items designed to measure global subjective happiness. Our participants indicated the degree of accuracy of each item in describing themselves on a 7-point scale. Their SHS score was the mean of the four items. Cronbach's alpha for this scale was .73.

Results

The Normative Structure of the Chinese NEO-PI-R

To identify the optimal number of factors underlying the Chinese NEO-PI-R in Dataset 1 ($N = 913$),[2] we used the Hull method (Lorenzo-Seva et al., 2011) and the fit indices for one- to eight-factor solutions estimated using exploratory factor analysis (EFA). Both analyses showed the optimal solutions to be five to seven factors (see Appendices 3 and 4 in the online supplement).[3] We examined each solution and found that the five-factor solution yielded the most interpretable results.

To examine the factor structure of the Chinese NEO-PI-R, we conducted ESEM in Mplus using robust maximum likelihood (MLR) estimation and oblique geomin rotation (Marsh et al., 2009). The goodness of fit of the models was evaluated using the comparative fit index (CFI), root mean square error of approximation (RMSEA), and Tucker-Lewis index (TLI). CFI or TLI values greater than .90 and RMSEA values below .08 reflect an acceptable fit (Marsh et al., 2004). The five-factor ESEM solution is presented in Table 2. The fit statistics for the ESEM solution are reported in the upper part of Table 3. The ESEM solution showed an adequate fit to the data (CFI = .911, TLI = .869, RMSEA = .061). Although TLI was slightly lower than the criterion for acceptable fit, this result (TLI < .90) is consistent with previous findings (Furnham et al., 2013). Therefore, ESEM provided clear support for the five-factor structure of the Chinese NEO-PI-R.

As shown in Table 2, the factor solution for the Chinese NEO-PI-R showed that all of the facets of Openness to Experience and Conscientiousness had their highest loadings on their target factors. All of the N facets except N5: Impulsiveness had their highest loadings on Neuroticism. N5 had nontrivial negative loadings on both Agreeableness and Conscientiousness. All of the E facets except E3: Assertiveness had their highest loadings on Extraversion, with E3 having a higher loading on Agreeableness than Extraversion. Three of the six Agreeableness facets (A1: Trust, A3: Altruism, and A6: Tender-Mindedness) had their highest loadings on Extraversion, implying that factor A was best conceptualized as a hybrid of E and A. This finding echoes Wiggins' (1979) description of interpersonal traits that fall between the orthogonal factors of E and A (see McCrae, Zonderman et al., 1996).

To evaluate the replicability of the Chinese NEO-PI-R using the American normative structure, we subjected the Chinese factor solution to an orthogonal Procrustes rotation using the normative factor values provided by Costa and McCrae (1992) as the target matrix and calculated congruence coefficients for each factor. These coefficients indicate the degree to which the solution obtained matches the (American) target matrix (McCrae et al., 1998), with a congruence coefficient of .85 or above considered evidence of factor replication (Lorenzo-Seva & ten Berge, 2006). As shown in the first row of Table 4, the factor congruence coefficients ranged from .92 (O) to .96 (N and C), indicating that the factor structure of the Chinese NEO-PI-R strongly resembled the American normative factor structure

Table 2. Factor Structure for the Chinese NEO-PI-R.

Facet	α	ESEM solution N	E	O	A	C
Neuroticism (N) facets						
N1: Anxiety	.741	.82	−.05	.03	−.01	.04
N2: Angry Hostility	.729	.63	−.06	−.06	−.51	−.06
N3: Depression	.808	.81	−.13	.06	.06	−.07
N4: Self-Consciousness	.666	.67	−.17	−.06	.11	−.04
N5: Impulsiveness	.649	.41	.15	.07	−.41	−.43
N6: Vulnerability	.776	.68	.06	−.17	.02	−.29
Extraversion (E) facets						
E1: Warmth	.717	.00	.75	.15	.08	.11
E2: Gregariousness	.718	−.08	.67	−.08	−.09	−.09
E3: Assertiveness	.745	−.19	.29	.13	−.51	.25
E4: Activity	.562	.00	.42	.03	−.40	.36
E5: Excitement-Seeking	.567	−.11	.27	.21	−.20	−.19
E6: Positive Emotions	.729	−.28	.58	.11	−.12	−.12
Openness to Experience (O) facets						
O1: Fantasy	.698	.00	−.01	.58	.00	−.28
O2: Aesthetics	.694	.13	.00	.65	.13	.01
O3: Feelings	.662	.29	.26	.58	−.12	.11
O4: Actions	.528	−.25	.15	.29	.03	−.09
O5: Ideas	.823	−.14	−.16	.65	−.07	.10
O6: Values	.334	−.13	.03	.28	.23	−.03
Agreeableness (A) facets						
A1: Trust	.773	−.11	.49	−.03	.38	.02
A2: Straightforwardness	.726	.02	.24	−.14	.52	.02
A3: Altruism	.636	.01	.52	.21	.40	.11
A4: Compliance	.552	−.03	.15	−.02	.66	.02
A5: Modesty	.717	.37	−.08	−.21	.47	−.10
A6: Tender-Mindedness	.509	.23	.40	.20	.23	−.07
Conscientiousness (C) facets						
C1: Competence	.658	−.29	.03	.28	−.28	.51
C2: Order	.737	.07	−.04	−.06	−.07	.67
C3: Dutifulness	.554	.02	.08	.05	.20	.67
C4: Achievement Striving	.758	−.03	.14	.11	−.28	.71
C5: Self-Discipline	.771	−.14	.04	.05	.00	.75
C6: Deliberation	.760	−.11	−.09	.07	.21	.61

Note. $N = 913$. α = Coefficient alpha. Coefficient alpha values are .921 for N, .870 for E, .841 for O, .842 for A, and .911 for C. Loadings equal to or greater than |.40| are presented underlined. NEO-PI-R = NEO Personality Inventory-Revised; ESEM = exploratory structural equation modeling.

(see McCrae, Zonderman et al., 1996; McCrae et al., 1998). The factor loadings of the Chinese NEO-PI-R were then set as the Chinese normative structure and used subsequently to test the structural properties of the Chinese NEO-PI-3.

The Psychometric Properties of the Chinese NEO-PI-3

Scale Analyses. Using data from Dataset 2 and Dataset 3: Time 1 ($N = 702$), we examined the psychometric properties of the G4 scales. Table 5 shows the coefficient alpha values for the Chinese NEO-PI-3. The internal consistency of these scales was comparable to, and even slightly better than, that found for our Chinese NEO-PI-R data (see Table 2): the median coefficient alpha values increased from .870 in the NEO-PI-R to .877 in the NEO-PI-3 for domain scores and from .717 to .724 for facet scores. Of the 30 facets, seven scales (E4, E5, O4, O6, A4, A6, and C3) had an alpha of less than .60 in the NEO-PI-R. The NEO-PI-3 showed marked improvement in two scales, A6: Tender-Mindedness ($\alpha = .655$) and C3: Dutifulness ($\alpha = .601$). The median coefficient alpha values across the 19 facets that included replacement items increased from .666 in the NEO-PI-R to .675 in the NEO-PI-3.

The means and standard deviations of the NEO-PI-3 scales are presented separately by gender in the Appendix.

Table 3. Summary of Model Fit Statistics in Exploratory Structure Equation Modeling.

Model	χ^2	*df*	RMSEA	CFI	TLI
Overall models					
NEO-PI-R (*N* = 913)	1,294.64	295	.061	.911	.869
NEO-PI-3 (*N* = 702)	1,040.45	295	.060	.906	.861
Models of language invariance					
1. Configural	2,160.70	1,435	.041	.937	.922
2. Metric (loadings)	2,429.19	1,560	.043	.924	.914
2p. Partial metric (loadings)[a]	2,372.00	1,546	.042	.928	.918
3. Strong (loadings, intercepts)[a]	2,683.48	1,571	.049	.903	.891
3p. Partial strong (loadings, intercepts)[a,b]	2,480.80	1,567	.044	.920	.910
4. Strict (loadings, intercepts, uniquenesses)[a,b]	2,592.84	1,597	.046	.913	.904
5. Manifest mean (loadings, intercepts, uniquenesses, factor means)[a,b]	2,796.87	1,602	.050	.896	.885

Note. χ^2 = chi-square fit statistics; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index; NEO-PI-R = NEO Personality Inventory-Revised; NEO-PI-3 = NEO Personality Inventory-3.
[a]Non-invariant loadings of N1: Anxiety and A2: Straightforwardness across languages. [b]Non-invariant intercepts of O4: Actions, A2: Straightforwardness, A4: Compliance, and C1: Competence across languages.

Table 4. Congruence Coefficients for the Factors in the Chinese NEO Scales.

Factor congruence	N	E	O	A	C
Chinese normative NEO-PI-R (*N* = 913)					
With the American normative NEO-PI-R structure[a]	.96	.95	.92	.94	.96
Chinese NEO-PI-3 (*N* = 702)					
With the Chinese normative NEO-PI-R structure	.99	.99	.99	.98	.99
With the American normative NEO-PI-R structure	.95	.93	.93	.94	.97

Note. NEO-PI-R = NEO Personality Inventory-Revised; NEO-PI-3 = NEO Personality Inventory-3; N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness.
[a]*N* = 1,000 (Costa & McCrae, 1992); the factor solution was obtained using a Varimax-rotated principal component analysis.

The pattern of these mean scores was similar to that found in our Chinese NEO-PI-R data (Dataset 1). For example, women scored significantly higher on N and A in both datasets (see Costa et al., 2001). Of the 13 significant gender differences found in the NEO-PI-R facet scales, 12 were replicated in our NEO-PI-3 data.

Factor Structure. Using data from the Chinese NEO-PI-3 (*N* = 702), the five-factor solution was examined using ESEM. The ESEM solution showed an adequate fit (CFI = .906, TLI = .861, RMSEA = .060; see the upper part of Table 3). As shown in Table 5, all of the facets had their highest loadings on the intended factors, providing support for the five-factor model. Notably, in contrast to the Chinese NEO-PI-R (see Table 2), in which three of the six A facets loaded most highly on E, all six facets in the Chinese NEO-PI-3 loaded most highly on A. This factor structure was therefore retained as a baseline model for further analyses of measurement invariance across groups.[4]

To test whether the Chinese and American normative structures could be recovered in our NEO-PI-3 data, we subjected the Chinese NEO-PI-3 factor solution to an orthogonal Procrustes rotation using each normative structure as the target matrix and calculated congruence coefficients for the five factors. First, the Chinese normative structure (Dataset 1) served as the target matrix. As shown in the second row of Table 4, the factor congruence coefficients were all above .97, supporting strong factor replicability. Next, we changed the target matrix to the American normative structure (Costa & McCrae, 1992). As shown in the final row of Table 4, the factor congruence coefficients were all above .92. In summary, the five factors were successfully recovered in our Chinese NEO-PI-3 data, and these factors were structurally similar to both Chinese and American normative structures. As anticipated, the congruence coefficients suggest that our Chinese NEO-PI-3 data had a factor structure more similar to the Chinese than to the American normative structure.

Equivalence of the Chinese Translation. We examined the cross-language validity of the NEO-PI-3 scales using data from Dataset 2 (*N* = 299).[5] A subsample of 149 participants (74 women) completed the test in Chinese and the retest in English, while the remaining 150 participants (72 women)

Table 5. Factor Structure for the Chinese NEO-PI-3.

		ESEM solution				
Facet	α	N	E	O	A	C
Neuroticism (N) facets						
N1: Anxiety	.704	.76	−.13	.03	.05	.01
N2: Angry Hostility	.759	.69	.04	−.06	−.42	−.01
N3: Depression	.814	.73	−.25	.02	.01	−.05
N4: Self-Consciousness	.725	.59	−.38	−.08	.11	−.07
N5: Impulsiveness	.632	.62	.26	.04	−.24	−.23
N6: Vulnerability	.762	.68	.04	−.17	.04	−.30
Extraversion (E) facets						
E1: Warmth	.724	−.01	.65	.18	.31	.10
E2: Gregariousness	.777	−.02	.71	−.05	.03	−.03
E3: Assertiveness	.668	−.16	.41	.14	−.33	.34
E4: Activity	.565	.07	.50	.02	−.12	.40
E5: Excitement-Seeking	.518	.01	.36	.21	−.06	−.09
E6: Positive Emotions	.803	−.23	.61	.13	.17	−.12
Openness to Experience (O) facets						
O1: Fantasy	.697	.00	.04	.50	−.02	−.29
O2: Aesthetics	.743	.09	−.07	.62	.11	−.06
O3: Feelings	.647	.41	.23	.54	.01	.16
O4: Actions	.584	−.21	.23	.31	.01	−.05
O5: Ideas	.822	−.12	−.27	.70	−.09	.08
O6: Values	.401	−.16	−.12	.35	.15	−.14
Agreeableness (A) facets						
A1: Trust	.799	−.10	.40	−.01	.44	−.04
A2: Straightforwardness	.745	.00	.02	−.14	.60	.05
A3: Altruism	.709	.02	.30	.19	.62	.11
A4: Compliance	.532	−.17	−.03	−.05	.65	−.07
A5: Modesty	.748	.28	−.30	−.17	.35	−.10
A6: Tender-Mindedness	.655	.23	.20	.23	.43	.10
Conscientiousness (C) facets						
C1: Competence	.675	−.30	.11	.22	−.15	.53
C2: Order	.794	−.03	−.12	−.04	.00	.62
C3: Dutifulness	.601	.05	−.03	.06	.31	.68
C4: Achievement Striving	.735	.04	.11	.06	−.15	.72
C5: Self-Discipline	.764	−.17	.10	.01	.04	.71
C6: Deliberation	.748	−.18	−.26	.04	.24	.55

Note. $N = 702$. α = Coefficient alpha. Coefficient alpha values for the domains are .923 for N, .877 for E, .840 for O, .851 for A, and .907 for C. Loadings equal to or greater than |.40| are presented underlined. NEO-PI-3 = NEO Personality Inventory-3; ESEM = exploratory structural equation modeling.

completed the test in English and the retest in Chinese. As there were no order effects, we pooled the data from these two groups.[6]

To assess the psychometric equivalence of the translations (Butcher & Han, 1996), we tested the measurement invariance across the two language versions (Chinese vs. English) using multiple indicator growth modeling. Correlated uniqueness between identical facets in the Chinese and English versions was used because the responses to the English NEO-PI-3 and its Chinese version were provided by the same group of participants (see the Mplus syntax in Appendices 6 to 8 of the online supplemental materials). In general, when comparing two nested models, the more constrained model is supported when CFI decreases by less than .010 and RMSEA increases by less than .015 (Chen, 2007; Cheung & Rensvold, 2002). The lower part of Table 3 shows the test results for a sequence of invariant models (Marsh et al., 2009; Meredith, 1993).

The first model we tested (Model 1) was a configural invariant model, a type of model with no invariance. Model 1 showed a good fit to the data (CFI = .937, TLI = .922, RMSEA = .041), thereby supporting configural invariance. The metric invariant model (Model 2), which

was a model with factor loadings constrained to be invariant across languages, showed an acceptable fit (CFI = .924, TLI = .914, RMSEA = .043) but resulted in a significant decline in fit relative to Model 1 (ΔCFI = −.013, ΔRMSEA = .002). We, therefore, explored partial metric invariance (Model 2p) by allowing the factor loadings of facets N1: Anxiety and A2: Straightforwardness (based on the highest modification indices) to be non-invariant across languages. We found that the cross-loading of N1 on Conscientiousness was trivial in Chinese (.11) but modest in English (.19), while the loading of A2 on Agreeableness was much greater in Chinese (.62) than in English (.48). There was also a trivial negative loading of A2 (−.11) on Neuroticism in English but a modest positive loading of A2 (.17) on Conscientiousness in Chinese. The fit of Model 2p was acceptable (CFI = .928, TLI = .918, RMSEA = .042), as was the change in fit (ΔCFI = −.009, ΔRMSEA = .001).[7]

Building on the partial metric invariant model (Model 2p), a strong invariant model (Model 3) was then tested by constraining the facet intercepts to be equal. Strong invariance would imply that differences in facet mean scores between language versions can be explained by differences at the latent factor level. This test resulted in a poorer fit (CFI = .903, TLI = .891, RMSEA = .049), with a considerable worsening of fit relative to Model 2p (ΔCFI = −.025, ΔRMSEA = .007). These results suggest that there might be differential item functioning in the two languages. Because the invariance of facet-level intercepts was crucial for testing cross-language differences in latent factor means, we moved to a partially strong invariance model (Model 3p). Allowing four facets (O4, A2, A4, and C1) to be non-invariant across languages substantially improved the model fit (CFI = .920, TLI = .910, RMSEA = .044) and the change in fit (ΔCFI = −.008, ΔRMSEA = .002). In the four facets with non-invariant mean scores, the participants tended to report lower scores in Chinese than in English for O4: Actions and C1: Competence and higher scores for A2: Straightforwardness and A4: Compliance. These results provided support for the partial strong invariance model with invariance in 26 of the 30 facet intercepts.

A test of strict invariance (Model 4) was then performed to constrain facet uniqueness to be invariant across languages (i.e., invariance in measurement errors). Model 4 showed an acceptable fit to the data (CFI = .913, TLI = .904, RMSEA = .046). The change in fit relative to Model 3p was also within a reasonable range (ΔCFI = −.007, ΔRMSEA = −.006). This achievement of strict invariance indicated equivalent measurement precision across the two languages. Building on the strict invariance model, a test of manifest mean invariance (Model 5) was conducted by constraining the factor means to be invariant across the two languages. This model considerably worsened the fit relative to Model 4 (ΔCFI = −.017, ΔRMSEA = .004), suggesting non-invariance in factor means across the two languages. We therefore deemed the strict invariant model (Model 4) to be the best-fitting model.

To test the difference in factor means across English and Chinese, we examined the strict invariance model (Model 4) in which the factor means were set to 0 in the Chinese version and freely estimated in the English version. Three of the five factors had lower means (Neuroticism: −.21, SE = .04, $Z = -4.97$, $p < .001$; Agreeableness: −.55, SE = .06, $Z = -8.60$, $p < .001$; Conscientiousness: −.21, SE = .04, $Z = -5.77$, $p < .001$) in the English version than in the Chinese version. As shown in Table 6, the factor solution for the Chinese version was very similar to that for the English version.

Reliability Estimates. We examined the test-retest reliability of the NEO-PI-3 in the two languages over a span of 2 weeks using Dataset 2 (N = 299). The third column of Table 7 presents the 2-week test-retest correlations for the cross-language (r_{ce}) sample. The correlations for the five domains exceeded .80, while those for the facet scales ranged from .60 (N5: Impulsiveness) to .85 (E2: Gregariousness) with a median of .74.[8] Of the 19 modified facets, 9 showed improvements in their cross-language coefficients. These results are comparable to the cross-language study of the NEO-PI-R in which McCrae et al. (1998) administered Chinese and English versions of the instrument to a small group of bilingual Hong Kong students (N = 81; median coefficient = .77).

We examined the test-retest reliability of the Chinese version of the NEO-PI-3 using Dataset 3: Time 2 (N = 403). The results are shown in the final column of Table 7 (r_{cc}). The 2-week test-retest correlations for the five domain scores ranged from .89 (A) to .92 (C) and the facet correlations were all above .70 (median = .82). All of the NEO-PI-3 domain and facet scales attained high correlations similar to those of the NEO-PI-R (see also McCrae & Costa, 2010).

Correlates. To test the affective underpinnings of the Chinese version of the G4 scales, we used Dataset 3: Time 2 (N = 403). First, we used Yik's (2009) 12 CCMA scales, testing their circumplex structure. The model fit the data well: c^2 (40, N = 403) = 177.41; RMSEA = .09. The 12 scales were located close to the predicted values: Pleasant was fixed at 0°, Activated was 91° away, Unpleasant was 169° away, and Deactivated was 294° away. The communality indices ranged from .78 to .99.

Using the cosine wave method proposed by Yik et al. (2011), we placed external variables one by one within the CCMA (see Yik, 2009) by fitting a cosine function to each series of 12 correlation coefficients between an external variable (e.g., the N domain score) and the CCMA scores. The circumplex model provides a powerful prediction: the 12 correlations form a cosine curve. In the cosine wave method, the fit of the correlation pattern to a cosine wave is

Table 6. Standardized Loading Estimates for the Chinese and English NEO-PI-3 in the Final Strict Invariant Model.

	Chinese					English				
Facet	N	E	O	A	C	N	E	O	A	C
Neuroticism (N) facets										
N1: Anxiety	.76	−.06	.06	.02	.11	.79	−.16	.00	.04	.19
N2: Angry Hostility	.68	−.12	−.03	−.46	−.03	.67	−.13	−.02	−.51	−.03
N3: Depression	.72	−.20	.03	.06	−.03	.69	−.21	.03	.06	−.03
N4: Self-Consciousness	.69	−.30	.04	.06	.05	.66	−.32	.04	.06	.05
N5: Impulsiveness	.58	.12	−.03	−.35	−.33	.54	.13	−.02	−.38	−.33
N6: Vulnerability	.73	.07	−.15	−.01	−.19	.69	.08	−.14	−.01	−.19
Extraversion (E) facets										
E1: Warmth	.06	.73	.19	.05	.14	.05	.71	.16	.05	.13
E2: Gregariousness	−.05	.66	−.15	−.15	−.05	−.05	.64	−.13	−.15	−.04
E3: Assertiveness	−.21	.24	.09	−.46	.24	−.19	.23	.07	−.45	.22
E4: Activity	.01	.40	−.08	−.38	.33	.01	.38	−.06	−.37	.30
E5: Excitement-Seeking	−.09	.37	.01	−.16	−.13	−.08	.36	.01	−.16	−.12
E6: Positive Emotions	−.15	.64	.15	−.02	−.06	−.14	.63	.13	−.02	−.06
Openness to Experience (O) facets										
O1: Fantasy	−.01	.09	.58	−.07	−.34	−.01	.09	.50	−.07	−.33
O2: Aesthetics	.10	.00	.56	.16	−.15	.09	.00	.48	.17	−.14
O3: Feelings	.40	.30	.42	−.17	.13	.35	.29	.36	−.17	.12
O4: Actions	−.24	.27	.21	.06	−.13	−.22	.26	.18	.06	−.12
O5: Ideas	−.16	−.27	.77	−.02	.04	−.15	−.29	.72	−.02	.04
O6: Values	−.05	.05	.33	.26	−.07	−.05	.05	.27	.26	−.06
Agreeableness (A) facets										
A1: Trust	−.12	.60	−.05	.33	−.02	−.11	.61	−.05	.34	−.02
A2: Straightforwardness	.08	.35	−.14	.62	.17	−.11	.27	−.13	.48	.02
A3: Altruism	.02	.58	.11	.44	.22	.02	.59	.09	.46	.21
A4: Compliance	−.13	.26	−.05	.57	.13	−.12	.26	−.04	.59	.12
A5: Modesty	.21	−.11	−.25	.46	−.07	.19	−.11	−.21	.46	−.06
A6: Tender-Mindedness	.20	.40	.17	.30	.13	.18	.39	.14	.30	.12
Conscientiousness (C) facets										
C1: Competence	−.32	.02	.13	−.20	.56	−.30	.02	.12	−.20	.54
C2: Order	−.02	−.07	−.15	−.04	.59	−.02	−.07	−.12	−.04	.55
C3: Dutifulness	.14	.15	.09	.15	.75	.13	.15	.08	.16	.73
C4: Achievement Striving	.01	.02	.03	−.27	.67	.01	.02	.03	−.27	.62
C5: Self-Discipline	−.20	.08	−.02	−.01	.70	−.19	.08	−.02	−.01	.68
C6: Deliberation	−.11	−.06	−.03	.19	.70	−.10	−.06	−.03	.20	.68

Note. $N = 299$. Loadings equal to or greater than |.40| are presented underlined. NEO-PI-3 = NEO Personality Inventory-3.

indicated by the variance accounted for (VAF). The magnitude of the relationship of the external variable to the entire circumplex (viz. affect) is indicated by r-max (r_{max}) and the location within the circumplex at which an external variable falls is indicated by a-hat (â). An r_{max} of ≥ .15 is required to conclude that an external variable is related to affect (see Yik, 2009; Yik et al., 2011).

We began by fitting the cosine function to each domain scale. Of the five domains, N, E, and C reached the hurdle of $r_{max} = .15$ (VAF: 88% to 98%), indicating that these three domains were substantially related to the CCMA. Neuroticism fell at 175°, Extraversion fell at 28°, and Conscientiousness fell at 355°. We then repeated the preceding analysis for each of the 30 facet scores. As shown in Table 8, 11 facets passed the hurdle of $r_{max} = .15$; the mean VAF value was 95% (range = 84%–99%). More than one third of the facet scales (six in N, two in E, one in A, and two in C) were strongly related to affect. Figure 1 shows a graph of the results.

To provide preliminary evidence of concurrent validity, we correlated the two well-being measures with the NEO-PI-3 domains. SWLS was found to be significantly related to N, E, O, A, and C (rs = −.41, .41, .12, .24, and .27, respectively; all ps $< .01$ except for O, in which $p < .05$).

Table 7. Two-Week Retest Coefficients of the Chinese NEO-PI-R and NEO-PI-3 Scales.

	NEO-PI-R		NEO-PI-3	
	McCrae et al. (1998)		*Dataset 2*	*Dataset 3*
	$n = 81$	$n = 41$	$n = 299$	$n = 403$
Scale	r_{ce}	r_{cc}	r_{ce}	r_{cc}
Domain (48 items each)				
N: Neuroticism	.90	.88	.86	.90
E: Extraversion	.90	.93	.89	.91
O: Openness to Experience	.86	.89	.82	.90
A: Agreeableness	.85	.92	.82	.89
C: Conscientiousness	.89	.95	.91	.92
Mdn	.89	.92	.86	.90
Facet (8 items each)				
N1: Anxiety[a]	.83	.82	.72	.80
N2: Angry Hostility	.73	.87	.67	.83
N3: Depression	.82	.76	.76	.84
N4: Self-Consciousness[a]	.70	.75	.74	.83
N5: Impulsiveness[a]	.72	.80	.60	.76
N6: Vulnerability	.86	.84	.76	.85
E1: Warmth	.86	.83	.80	.86
E2: Gregariousness	.85	.91	.85	.86
E3: Assertiveness[a]	.82	.92	.81	.85
E4: Activity[a]	.71	.72	.70	.76
E5: Excitement-Seeking[a]	.80	.85	.74	.82
E6: Positive Emotions[a]	.80	.89	.77	.85
O1: Fantasy	.79	.85	.67	.81
O2: Aesthetics[a]	.79	.87	.81	.86
O3: Feelings	.58	.75	.64	.78
O4: Actions[a]	.69	.81	.70	.71
O5: Ideas	.86	.88	.79	.89
O6: Values[a]	.46	.77	.61	.71
A1: Trust[a]	.79	.80	.76	.81
A2: Straightforwardness[a]	.69	.82	.68	.84
A3: Altruism	.76	.81	.72	.82
A4: Compliance[a]	.56	.77	.64	.77
A5: Modesty[a]	.74	.84	.79	.80
A6: Tender-Mindedness[a]	.51	.63	.65	.81
C1: Competence[a]	.75	.81	.73	.83
C2: Order[a]	.77	.87	.81	.86
C3: Dutifulness[a]	.73	.77	.71	.78
C4: Achievement Striving[a]	.79	.88	.81	.85
C5: Self-Discipline	.75	.88	.79	.83
C6: Deliberation	.80	.87	.75	.80
Mdn	.77 (.51–.86)	.83 (.63–.92)	.74 (.60–.85)	.82 (.71–.89)

Note. r_{ce} = Equivalence correlation for the Chinese and English versions; r_{cc} = Retest correlation for the Chinese version. All correlations were significant at $p < .05$. NEO-PI-R = NEO Personality Inventory-Revised; NEO-PI-3 = NEO Personality Inventory-3.
[a]Facet changed in the NEO-PI-3.

Table 8. Affective Core of the NEO-PI-3 and Well-Being Scales.

Scale	α	Parameter estimates of the cosine wave method		
		â	r_{max}	VAF (%)
N: Neuroticism	.925	175°	.33	98**
N1: Anxiety	.781	179°	.26	97**
N2: Angry hostility	.729	164°	.28	99**
N3: Depression	.840	179°	.33	97**
N4: Self-Consciousness	.746	185°	.23	96**
N5: Impulsiveness	.652	166°	.18	93**
N6: Vulnerability	.755	172°	.25	97**
E: Extraversion	.879	28°	.18	92**
E1: Warmth	.760	356°	.13	88**
E2: Gregariousness	.770	26°	.09	88**
E3: Assertiveness	.694	46°	.08	96**
E4: Activity	.525	50°	.15	96**
E5: Excitement-Seeking	.572	59°	.08	66**
E6: Positive Emotions	.821	20°	.23	84**
O: Openness to Experience	.849	41°	.08	79**
O1: Fantasy	.701	86°	.04	62**
O2: Aesthetics	.781	53°	.08	81**
O3: Feelings	.706	129°	.05	76**
O4: Actions	.549	51°	.11	89**
O5: Ideas	.860	2°	.11	74**
O6: Values	.430	297°	.08	85**
A: Agreeableness	.858	310°	.13	89**
A1: Trust	.794	343°	.19	95**
A2: Straightforwardness	.781	287°	.08	75**
A3: Altruism	.678	354°	.10	88**
A4: Compliance	.583	324°	.12	88**
A5: Modesty	.762	211°	.13	94**
A6: Tender-Mindedness	.671	325°	.11	83**
C: Conscientiousness	.909	355°	.15	88**
C1: Competence	.669	356°	.19	96**
C2: Order	.781	46°	.02	33
C3: Dutifulness	.616	326°	.14	92**
C4: Achievement Striving	.737	17°	.13	91**
C5: Self-Discipline	.767	351°	.17	93**
C6: Deliberation	.763	349°	.09	76**
Well-being				
Satisfaction with Life	.878	355°	.32	96**
Subjective Happiness	.731	6°	.28	92**

Note. $N = 403$. α = Coefficient alpha. â (a-hat) is the estimated angle of the external variable within the CCMA structure; r_{max} (r-max) is the maximum correlation between the external variable and a vector within the CCMA at the angle â; VAF (variance accounted for) is the amount of variance explained by the cosine function. NEO-PI-3 = NEO Personality Inventory-3.
*$p < .05$. **$p < .01$.

SHS was also significantly related to N, E, O, A, and C (rs = −.48, .59, .16, .17, and .22, respectively; all ps < .01). The two well-being measures were more strongly related to N and E than to the other domains.

To provide an alternative method for mapping the relationship between the FFM and the well-being measures, we again used the cosine wave method to place the well-being measures within the CCMA. SHS fell at 6°, SWLS fell at 355°, and both were significantly related to affect. As shown in Figure 1, both of these measures fell close to the horizontal axis of the CCMA and were approximately 180° away from the N cluster (echoed by their significantly negative correlations with N).

Discussion

The results of this study provide encouragement to those who are interested in using an adequate measure to capture

personality in Chinese communities that also allows them to compare personality traits across the globe. To test the factor structure of the fourth generation of the NEO scales, we adopted ESEM, multiple indicator growth modeling, and multi-group comparisons to test measurement invariance across languages and gender groups. The tests of language invariance revealed reasonable partial metric and partial strong measurement invariance, indicating that the observed facet scores and variances can be meaningfully compared across languages. Non-equivalence was observed for only two loadings (N1: Anxiety and A2: Straightforwardness) and four intercepts (O4: Actions, A2: Straightforwardness, A4: Compliance, and C1: Competence) of the facet scores. The data also supported strict invariance, implying equivalent precision of the English and Chinese versions of the NEO-PI-3. Taken together, these results provide strong support for the cross-language generalizability of the NEO-PI-3 factor structure (Meredith, 1993).

The present study constitutes the first assessment of test-retest reliability for the Chinese NEO-PI-3. The 35 (5 domains and 30 facets) test-retest correlations were statistically significant, indicating stability of the scores over 2 weeks (.71 to .92). McCrae et al. (2011) demonstrated that test-retest reliability, not internal consistency, is the strongest predictor of a scale's validity. The test-retest coefficients reported for our NEO-PI-3 scores are adequate and provide support for the use of this instrument with Chinese participants. The NEO-PI-3 scales performed similarly to their immediate precedent, the NEO-PI-R. They maintained high levels of cross-language equivalence and test-retest reliability alongside improved internal consistency and readability. However, we suspect that using samples of university students may have limited our ability to capture the improvements in the psychometric properties of the G4 scales. Future research should target a wider range of populations to test whether the new scales are appropriate for those with a reading level as low as Grade 5, which was one of the key drivers behind the release of the NEO-PI-3 scales (McCrae & Costa, 2010; McCrae, Costa, & Martin, 2005).

Personality and Language

In the present research, we used a bilingual test-retest design to examine the effects of linguistic and cultural contingencies on the responses while eliminating the variance due to individual differences (Butcher et al., 2006). Despite the high degree of congruence in the five-factor model across the two languages, there were several observable differences depending on whether our bilingual participants were responding in Chinese or English. The non-invariance of the factor loadings of N1: Anxiety and A2: Straightforwardness echoes previous cross-cultural analyses of differential item functioning in the NEO scales (see Lui et al., 2020; Rollock & Lui, 2016). Our bilingual participants reported higher Neuroticism, Agreeableness, and Conscientiousness when responding in Chinese than when responding in English (cf. Chen, Lam et al., 2014). The mean differences between the two language versions could be attributed to translation, reference group effects (Heine et al., 2002), or cultural frame switching (Hong et al., 2000).

Focusing on the functional features of language, Chen and Bond (2010) found that when a Chinese questionnaire was used, the language oriented the respondents to the thoughts and ideas of Chinese culture. As such, when responding in Chinese, their Hong Kong Chinese bilingual respondents were influenced by prototypical characteristics of Chinese speakers, resulting in higher scores for Neuroticism but lower scores for Extraversion, Openness to Experience, Agreeableness, and Conscientiousness than when they responded in English (see also Chen, Benet-Martínez, & Ng, 2014). Consistent with these previous findings, our Hong Kong bilingual participants scored higher on Neuroticism in Chinese than in English (cf. McCrae et al., 1998). However, our results for Agreeableness and Conscientiousness were not consistent with these findings. Perhaps our bilingual participants perceived themselves as more agreeable and conscientious than their Chinese peers (the reference group) or applied a lower standard in rating these personality dimensions when responding in Chinese (see McCrae et al., 1998).

To summarize, the differences between the English and Chinese versions suggest that multiple processes may have been involved. Future research is needed to provide a more systematic examination of the mechanisms involved in personality assessment across languages. For instance, we could examine the frames of reference used by bilingual participants when completing the questionnaire in each language.

Personality and Affect

To examine the correlations of the NEO scales with affect and well-being, we took an innovative approach by placing the 35 NEO scales and well-being measures within the CCMA (Yik, 2009). The CCMA scales capture the full spectrum of feelings that are hypothesized to be composed of varying proportions of valence and arousal. The cosine wave method showed that 14 of the 35 NEO scales and the two well-being measures were significantly related to affect. These external correlates clustered at either end of the valence (horizontal) axis of the circumplex model. The N domain and its facet scales clustered between 164° and 185°, indicating that they shared very similar affective underpinnings. People who are high on Neuroticism tend to experience unpleasant affect with a medium level of arousal (e.g., *unhappy, downhearted*). The well-being

measures we examined fell close to the C scales: people who are high on C (C1: Competence, C5: Self-Discipline) or are satisfied with their lives tend to experience pleasant affect with medium arousal (e.g., *content, carefree*). It is prudent to conclude that the horizontal axis of valence (with medium arousal) serves as an affective intersection plane with the FFM and well-being (see Yik, 2010b; Yik et al., 2002).

Tsai et al. (2006) argued that people of different cultural backgrounds crave different emotions. For example, compared with European Americans (who value high arousal positive affect, e.g., *excited*), Chinese people value low arousal positive affect (e.g., *calm*). These cultural differences are robust across age groups (Tsai et al., 2007). Therefore, feeling "calm" (pleasant affect with low arousal) may be highly valued in Chinese culture and also significantly related to subjective well-being. Contrary to the findings of Tsai et al. (2007), our results showed that people who are satisfied or happy with their lives tend to experience pleasant affect with medium but not low arousal (6°; 355°). In the future, this research should be replicated by including both current affect (what was measured in this study) and ideal affect (what people want to experience). The relationship between these two types of affect has yet to be tested and may relate differently to subjective well-being in Chinese culture.

People who are high on Extraversion or Subjective Happiness tend to experience pleasant affect with a moderately high level of arousal (e.g., *lively, overjoyed*). Interestingly, McCrae et al. (1998) found that regardless of where their Hong Kong Chinese participants were located, they were less extraverted than North Americans (see also McCrae, Terracciano, et al., 2005). These results suggest that Hong Kong Chinese may feel less happy and lively than people in other communities. This observation aligns well with a recent happiness survey in which Hong Kong was ranked 42nd and the United States was ranked 29th (Gallup International Association, 2021).

Alternative Measures of Personality

In the present study, we adopted an imposed etic design (Berry, 1969) in adopting the NEO scales for use in Chinese communities. Other examples of this imposed etic design include the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1951), the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1991), the Sixteen Personality Factor Questionnaire (Cattell & Mead, 2008), and the Big Five Inventory (Zhang et al., 2022). This style of research presupposes the universal or etic status of the concept of personality and has therefore been criticized for overlooking concepts that are indigenous to Chinese culture.

Some researchers have spearheaded an emic, or indigenous, approach by developing scales that assess Chinese-specific constructs. For example, Xia and Huang (2008) developed the Self-Supporting Personality Scale for Adolescent Students to measure the self-supporting personality trait, which enhances problem-solving in daily life and personal growth. Jiao et al. (2019) developed the Good Personality Lexical Rating Scale and the Evil Personality Lexical Rating Scale to assess moral character. Wang and Cui (2003) developed the Qingnian Zhongguo Personality Scale (QZPS) to measure the Chinese Big Seven: Extraversion, Kindness, Behavior Styles, Talents, Emotionality, Human Relationships, and Ways of Life.

These instruments and their associated models might result in scales that tap different aspects of personality. After all, personality is a complicated concept, and each instrument represents only an approximation of the full picture. Because of this, some researchers have championed a hybrid approach in which they deploy both emic and imposed etic measures to unravel the full complexities of Chinese personality. For instance, Yik and Bond (1993) used scales that tapped the Big Five traits, imported from English, alongside items selected from Chinese written materials. They found that eight factors represent an adequate description of Chinese personality. Cheung et al. (2008) used the NEO-FFI and the Revised CPAI-2 and mapped five factors, four of which resembled the FFM factors.

To further test the utility of the Chinese NEO-PI-3 in mapping personality, future research should aim to expand the nomological net of the scales: how are they related to other emic measures of personality and to outcome variables, such as job performance or happiness (see Yik, 2022)? These scales will provide a framework for the next decade of personality research in Chinese communities. If the FFM can be seen as a unified framework for personality research, perhaps these NEO-PI-3 scales can serve as a platform on which researchers can discuss their findings of stability, heritability, life outcomes, and cross-cultural similarities and differences.

Appendix

Means and Standard Deviations of the NEO-PI-R and NEO-PI-3 Scales.

		Chinese NEO-PI-R (Dataset 1)							Chinese NEO-PI-3 (Datasets 2 and 3)						
		Combined (N = 913)		Male (n = 330)		Female (n = 583)			Combined (N = 702)		Male (n = 334)		Female (n = 368)		
Scale	Range	M	(SD)	M	(SD)	M	(SD)	t diff.	M	(SD)	M	(SD)	M	(SD)	t diff.
Domain (48 items each)															
N	0 to 192	102.85	(21.66)	98.78	(22.08)	105.16	(21.10)	−4.32***	107.00	(21.69)	102.43	(21.12)	111.15	(21.38)	−5.42***
E	0 to 192	106.63	(17.47)	106.35	(17.61)	106.79	(17.41)		100.36	(18.17)	99.39	(18.38)	101.24	(17.96)	
O	0 to 192	112.21	(15.91)	112.75	(16.04)	111.90	(15.85)		114.38	(15.88)	114.78	(15.71)	114.02	(16.05)	
A	0 to 192	108.04	(15.17)	104.36	(15.22)	110.13	(14.75)	−5.61***	109.84	(16.01)	106.47	(15.99)	112.91	(15.42)	−5.43***
C	0 to 192	112.85	(19.31)	114.42	(18.66)	111.97	(19.63)		106.12	(19.17)	107.21	(19.51)	105.12	(18.83)	
Facet (8 items each)															
N1	0 to 32	18.99	(4.63)	17.94	(4.65)	19.58	(4.51)	−5.23***	20.20	(4.29)	19.39	(4.35)	20.94	(4.10)	−4.86***
N2	0 to 32	16.33	(4.69)	15.65	(4.76)	16.72	(4.61)	−3.31***	16.20	(4.85)	15.40	(4.81)	16.92	(4.77)	−4.21***
N3	0 to 32	17.34	(5.26)	16.94	(5.52)	17.57	(5.10)		18.88	(5.25)	18.29	(5.12)	19.42	(5.31)	−2.86**
N4	0 to 32	18.14	(4.31)	17.86	(4.34)	18.29	(4.28)		19.53	(4.68)	18.92	(4.58)	20.07	(4.71)	−3.28**
N5	0 to 32	16.69	(4.33)	16.42	(4.30)	16.84	(4.34)		16.27	(4.18)	15.64	(4.14)	16.85	(4.13)	−3.88***
N6	0 to 32	15.37	(4.58)	13.96	(4.48)	16.16	(4.45)	−7.16***	15.92	(4.46)	14.80	(4.43)	16.94	(4.25)	−6.54***
E1	0 to 32	21.60	(4.00)	21.35	(4.09)	21.74	(3.95)		20.99	(4.17)	20.58	(4.22)	21.36	(4.11)	−2.48*
E2	0 to 32	16.56	(4.49)	16.34	(4.36)	16.68	(4.56)		14.98	(4.95)	14.73	(5.23)	15.21	(4.68)	
E3	0 to 32	14.82	(4.60)	15.28	(4.66)	14.55	(4.55)	2.31*	13.79	(4.20)	14.22	(4.07)	13.41	(4.29)	2.57*
E4	0 to 32	17.03	(3.81)	17.02	(3.87)	17.03	(3.78)		15.76	(3.82)	15.56	(3.92)	15.95	(3.71)	
E5	0 to 32	17.65	(4.41)	17.93	(4.46)	17.50	(4.39)		16.61	(4.23)	16.63	(4.32)	16.58	(4.16)	
E6	0 to 32	18.98	(4.45)	18.42	(4.61)	19.30	(4.33)	−2.85**	18.22	(5.18)	17.67	(5.08)	18.72	(5.22)	−2.69**
O1	0 to 32	18.28	(4.65)	18.15	(4.51)	18.35	(4.73)		18.67	(4.64)	18.46	(4.63)	18.86	(4.64)	
O2	0 to 32	19.27	(4.73)	18.99	(4.86)	19.44	(4.64)		19.23	(5.12)	18.64	(5.19)	19.75	(5.01)	−2.88**
O3	0 to 32	20.68	(3.85)	20.30	(3.83)	20.90	(3.85)	−2.26*	20.63	(3.88)	20.22	(3.99)	21.00	(3.75)	−2.67**
O4	0 to 32	14.79	(3.47)	14.53	(3.62)	14.93	(3.38)		14.86	(3.53)	14.65	(3.71)	15.05	(3.35)	
O5	0 to 32	19.10	(5.49)	20.85	(5.30)	18.10	(5.36)	7.49***	20.31	(5.37)	21.89	(4.94)	18.89	(5.35)	7.69***
O6	0 to 32	20.09	(3.00)	19.93	(3.38)	20.18	(2.75)		20.69	(3.18)	20.93	(3.54)	20.48	(2.80)	
A1	0 to 32	18.82	(4.35)	18.12	(4.54)	19.21	(4.20)	−3.69***	16.87	(4.66)	16.35	(4.84)	17.35	(4.44)	−2.87**
A2	0 to 32	16.57	(4.67)	15.11	(4.68)	17.39	(4.46)	−7.29***	17.75	(4.97)	17.01	(5.08)	18.42	(4.77)	−3.78***
A3	0 to 32	19.85	(3.62)	19.58	(3.69)	20.00	(3.57)		20.44	(4.00)	19.94	(4.08)	20.90	(3.88)	−3.22**
A4	0 to 32	17.41	(3.86)	17.04	(4.00)	17.63	(3.77)	−2.22*	18.03	(3.81)	17.60	(3.76)	18.41	(3.81)	−2.85**
A5	0 to 32	15.30	(4.27)	14.60	(4.58)	15.70	(4.04)	−3.63***	16.03	(4.71)	15.57	(4.76)	16.45	(4.63)	−2.47*
A6	0 to 32	20.10	(3.30)	19.92	(3.56)	20.20	(3.15)		20.72	(3.80)	20.01	(4.07)	21.38	(3.41)	−4.81***
C1	0 to 32	17.87	(3.86)	18.88	(3.80)	17.30	(3.78)	6.06***	15.65	(3.84)	16.18	(3.73)	15.18	(3.88)	3.49***
C2	0 to 32	18.12	(4.78)	18.15	(4.64)	18.10	(4.87)		16.77	(5.27)	17.29	(5.44)	16.30	(5.07)	2.49*
C3	0 to 32	21.42	(3.42)	21.28	(3.43)	21.50	(3.41)		20.55	(3.51)	20.25	(3.60)	20.82	(3.40)	-2.16*
C4	0 to 32	19.35	(4.34)	19.78	(4.39)	19.10	(4.29)	2.29*	18.01	(4.40)	18.28	(4.44)	17.76	(4.36)	
C5	0 to 32	18.26	(4.56)	18.20	(4.44)	18.30	(4.63)		16.92	(4.67)	16.83	(4.67)	17.01	(4.67)	
C6	0 to 32	17.83	(4.53)	18.12	(4.25)	17.67	(4.67)		18.21	(4.39)	18.38	(4.56)	18.06	(4.23)	

Note. NEO-PI-R = NEO Personality Inventory-Revised; NEO-PI-3 = NEO Personality Inventory-3; N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness.
*$p < .05$. **$p < .01$. ***$p < .001$.

Acknowledgment

We thank Jeff McCrae, Dennis Li, Luna Li, and James Pho for their help in preparing this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of the paper was facilitated by the Hong Kong Research Grants Council's General Research Fund (Projects 16601818 and 16601921).

ORCID iDs

Michelle Yik https://orcid.org/0000-0003-0104-3662
Felity H. C. Kwok https://orcid.org/0000-0003-2179-720X
Shiang-yi Lin https://orcid.org/0000-0002-8001-7077

Supplemental Material

Supplemental material for this article is available online.

Notes

1. In Hong Kong, people primarily speak Cantonese and write using traditional Chinese characters.
2. Monte Carlo simulations were conducted to assess the required sample size and the precision of parameter estimation for factor analyses (Muthén & Muthén, 2002). Datasets of 5,000 replications with a sample size of 700 were generated using a five-factor ESEM model with 30 items. The results showed that our studies, with their sample sizes of over 700 (N = 913 for the NEO-PI-R and N = 702 for the NEO-PI-3) had sufficient power to detect the significance of the parameters in factor analyses (see Appendix 1 in the online supplement for more details).
3. To provide the fit statistics for models with different numbers of factors, we conducted EFA using oblique geomin rotation in Mplus. The five-factor solution achieved a marginally acceptable fit (CFI = .913, TLI = .872, RMSEA = .064; see the fit statistics in Appendix 4 of the online supplement).
4. To examine gender differences in the structure of the Chinese NEO-PI-3, multigroup ESEM was estimated in Mplus to test the invariance of the five-factor structure across men and women. The results provided support for measurement invariance: the CFI, TLI and RMSEA values were .897, .848 and .063 for configural invariance, .903, .881 and .056 for metric invariance, .896, .877 and .057 for partial strong invariance, and .887, .868 and .059 for latent mean invariance.
5. Monte Carlo simulations were carried out to evaluate the power of the sample size (N = 299) for testing measurement invariance across the two languages in the bilingual design. We tested the power with a sample size of 300 over 5,000 replications. This sample size was demonstrated to afford sufficient power to test language invariance (see Appendix 2 in the online supplement for more details).
6. Of the 70 possible t-tests, none were significant.
7. A test of the invariance of the factor variance-covariance was also conducted by comparing the partial metric invariant model (Model 2p) with a nested model in which the factor variance-covariance matrix was set to be invariant across languages. The test resulted in only slight changes in fit (ΔCFI = –.004, ΔRMSEA = .001), thereby supporting the invariance of the factor variance-covariance matrix across languages.
8. Supplementary analyses showed that the cross-language test-retest correlations were similar for those with high (> level 4) and low (< level 5) grades on a standard English language exam. Possible grade levels for this English test are, from lowest to highest: Unclassified, 1, 2, 3, 4, 5, 5*, and 5**.

References

Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., & Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. *Psychological Bulletin*, *146*(4), 279–323. https://doi.org/10.1037/bul0000226

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438. https://doi.org/10.1080/10705510903008204

Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology*, *4*(2), 119–128. https://doi.org/10.1080/00207596908247261

Butcher, J. N., & Han, K. (1996). Methods of establishing cross-cultural equivalence. In J. N. Butcher (Ed.), *International adaptations of the MMPI-2: Research and clinical applications* (pp. 44–63). University of Minnesota Press.

Butcher, J. N., Mosch, S. C., Tsai, J., & Nezami, E. (2006). Cross-cultural applications of the MMPI-2. In J. N. Butcher (Ed.), *MMPI-2: A practitioner's guide* (pp. 505–537). American Psychological Association. https://doi.org/10.1037/11287-018

Cattell, H. E. P., & Mead, A. D. (2008). The Sixteen Personality Factor Questionnaire (16PF). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Personality measurement and testing* (Vol. 2, pp. 135–159). SAGE. https://doi.org/10.4135/9781849200479.n7

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, P., Wu, J., Luo, L., Gao, H., Wang, M., Zou, X., Li, Y., Chen, G., Luo, H., Yu, L., Han, Y., Jia, F., & He, G. (2019). Population genetic analysis of modern and ancient DNA variations yields new insights into the formation, genetic structure, and phylogenetic relationship of Northern Han Chinese. *Frontiers in Genetics*, *10*, Article 1045. https://doi.org/10.3389/fgene.2019.01045

Chen, S. X., Benet-Martínez, V., & Ng, J. C. K. (2014). Does language affect personality perception? A functional approach to testing the Whorfian hypothesis. *Journal of Personality*, *82*(2), 130–143. https://doi.org/10.1111/jopy.12040

Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, *36*(11), 1514–1528. https://doi.org/10.1177/0146167210385360

Chen, S. X., Lam, B. C. P., Buchtel, E. E., & Bond, M. H. (2014). The conscientiousness paradox: Cultural mindset shapes competence perception. *European Journal of Personality*, *28*(5), 425–436. https://doi.org/10.1002/per.1923

Cheung, F. M., Cheung, S. F., Zhang, J., Leung, K., Leong, F., & Yeh, K. H. (2008). Relevance of Openness as a personality dimension in Chinese culture: Aspects of its cultural relevance. *Journal of Cross-Cultural Psychology*, *39*(1), 81–108. https://doi.org/10.1177/0022022107311968

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Costa, P. T., Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1993). Bullish on personality psychology. *The Psychologist*, *6*(7), 302–303.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. https://doi.org/10.1037/0022-3514.81.2.322

Dai, G., Han, K., Hu, H., & Colarelli, S. M. (2010). Cultural differences and measurement invariance of selection tools: A case of examining Chinese NEO PI-R conscientiousness scale. *Journal of Chinese Human Resource Management*, *1*(2), 95–114. https://doi.org/10.1108/20408001011117644

De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., Costa, P. T., Jr., & Collaborators of the Adolescent Personality Profiles of Cultures Project. (2009). Assessing the universal structure of personality in early adolescence: The NEO-PI-R and NEO-PI-3 in 24 cultures. *Assessment*, *16*(3), 301–311. https://doi.org/10.1177/1073191109333760

De Vries, R. E., Realo, A., & Allik, J. (2016). Using personality item characteristics to predict single-item internal reliability, retest reliability, and self-other agreement. *European Journal of Personality*, *30*(6), 618–636. https://doi.org/10.1002/per.2083

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2021). *Ethnologue: Languages of the world* (24th ed.). SIL International. https://www.ethnologue.com/

Eysenck, H. J., & Eysenck, S. B. G. (1991). *Manual of the Eysenck personality scales (adults)*. Hodder and Stoughton.

Fountoulakis, K. N., Siamouli, M., Moysidou, S., Pantoula, E., Moutou, K., Panagiotidis, P., Kemeridou, M., Mavridou, E., Loli, E., Batsiari, E., Preti, A., Tondo, L., Gonda, X., Mobayed, N., Akiskal, K., Akiskal, H., Costa, P. T., Jr., & McCrae, R. (2014). Standardization of the NEO-PI-3 in the Greek general population. *Annals of General Psychiatry*, *13*(1), Article 36. https://doi.org/10.1186/s12991-014-0036-9

Furnham, A., Guenole, N., Levine, S. Z., & Chamorro-Premuzic, T. (2013). The NEO Personality Inventory-Revised: Factor structure and gender invariance from exploratory structural equation modeling analyses in a high-stakes setting. *Assessment*, *20*(1), 14–23. https://doi.org/10.1177/1073191112448213

Gallup International Association. (2021, December 28). *Less hope and happiness in the world, 2021*. https://www.gallup-international.com/survey-results-and-news/survey-result/less-hope-and-happiness-in-the-world-2021

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

Hathaway, S. R., & McKinley, J. C. (1951). *The MMPI manual*. The Psychological Corporation.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903–918. https://doi.org/10.1037/0022-3514.82.6.903

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hong, Y.-y., Morris, M. W., Chiu, C.-y., & Benet-Martínez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, *55*(7), 709–720. https://doi.org/10.1037/0003-066X.55.7.709

Hrebícková, M. (2008). NEO Personality Inventory (NEO-PI-3) in diagnostics of children and adolescents. *Ceskoslovenska Psychologie*, *52*(5), 425–442.

Jiao, L. Y., Yang, Y., Xu, Y., Gao, S. Q., & Zhang, H. Y. (2019). Good and evil in Chinese culture: Personality structure and connotation. *Acta Psychologica Sinica*, *51*(10), 1128–1142. https://doi.org/10.3724/SP.J.1041.2019.01128

Källmen, H., Wennnberg, P., Andreasson, P., & Bergman, H. (2016). Psychometric properties of the Swedish version of the personality inventory NEO-PI-3. *International Journal of Psychology and Psychoanalysis*, *2*(1), 1–3. https://doi.org/10.23937/2572-4037.1510011

Kuppens, P., Tuerlinckx, F., Yik, M., Koval, P., Coosemans, J., Zeng, K. J., & Russell, J. A. (2017). The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of Personality*, *85*(4), 530–542. https://doi.org/10.1111/jopy.12258

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., Rubio-Codina, M., & Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science Advances*, *5*(7), Article eaaw5226. https://doi.org/10.1126/sciadv.aaw5226

Le Corff, Y., & Busque-Carrier, M. (2016). Structural validity of the NEO Personality Inventory 3 (NEO-PI-3) in a French-Canadian sample. *International Journal of Arts & Sciences*, *9*(3), 461–472.

Lorenzo-Seva, U., & ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340–364. http://dx.doi.org/10.1080/00273171.2011.564527

Lui, P. P., Samuel, D. B., Rollock, D., Leong, F. T. L., & Chang, E. C. (2020). Measurement invariance of the five factor model of personality: Facet-level analyses among Euro and Asian Americans. *Assessment*, *27*(5), 887–902. https://doi.org/10.1177/1073191119873978

Luo, J., & Dai, X. (2011). Meta-analysis of Big-five factor personality tests in China. *Chinese Journal of Clinical Psychology*, *19*(6), 740–752.

Lyubomirsky, S. (2008). *The how of happiness: A scientific approach to getting the life you want*. Penguin Press.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indices and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341. https://doi.org/10.1207/s15328007sem1103_2

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. Structural. *Equation Modeling*, *16*, 439–476. https://doi.org/10.1080/10705510903008220

McCrae, R. R., & Allik, J. (Eds.). (2002). *The Five-Factor Model of personality across cultures*. Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0763-5

McCrae, R. R., & Costa, P. T., Jr. (1983). Joint factors in self-reports and ratings: Neuroticism, Extraversion and Openness to Experience. *Personality and Individual Differences*, *4*(3), 245–255. https://doi.org/10.1016/0191-8869(83)90146-0

McCrae, R. R., & Costa, P. T., Jr. (1991). Adding Liebe und Arbeit: The full five-factor model and well-being. *Personality and Social Psychology Bulletin*, *17*(2), 227–232. https://doi.org/10.1177/014616729101700217

McCrae, R. R., & Costa, P. T., Jr. (2010). *NEO Inventories for the NEO Personality Inventory-3 (NEO-PI-3), NEO Five-Factor Inventory-3 (NEO-FFI-3), NEO Personality Inventory-Revised (NEO PI-R): Professional manual*. Psychological Assessment Resources.

McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, *84*(3), 261–270. https://doi.org/10.1207/s15327752jpa8403_05

McCrae, R. R., Costa, P. T., Jr., & Yik, M. (1996). Universal aspects of Chinese personality structure. In M. H. Bond (Ed.), *The handbook of Chinese psychology* (pp. 189–207). Oxford University Press.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, *15*(1), 28–50. https://doi.org/10.1177/1088868310366253

McCrae, R. R., & Terracciano, A., & 79 Members of the Personality Profiles of Cultures Project. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, *89*(3), 407–425. https://doi.org/10.1037/0022-3514.89.3.407

McCrae, R. R., Yik, M., Trapnell, P. D., Bond, M. H., & Paulhus, D. L. (1998). Interpreting personality profiles across cultures: Bilingual, acculturation, and peer rating studies of Chinese undergraduates. *Journal of Personality and Social Psychology*, *74*(4), 1041–1055. https://doi.org/10.1037/0022-3514.74.4.1041

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552–566. https://doi.org/10.1037/0022-3514.70.3.552

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543. https://doi.org/10.1007/BF02294825

Muthén, L. K., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. https://doi.org/10.1037/a0014996

Quy, G. S. (2011). *Adapting the NEO-PI-3 for a South African context: A pilot study using a South African student population* [Master's thesis, University of the Witwatersrand, Johannesburg]. Electronic Theses and Dissertations. http://hdl.handle.net/10539/9752

Rollock, D., & Lui, P. P. (2016). Measurement invariance and the five-factor model of personality: Asian international and Euro American cultural groups. *Assessment*, *23*(5), 571–587. https://doi.org/10.1177/1073191115590854

Siegler, I. C., & Brummett, B. H. (2000). Associations among NEO personality assessments and well-being at midlife: Facet-level analyses. *Psychology and Aging*, *15*(4), 710–714. https://doi.org/10.1037/0882-7974.15.4.710

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), 117–143. https://doi.org/10.1037/pspp0000096

Strickhouser, J. E., Zell, E., & Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychology*, *36*(8), 797–810. https://doi.org/10.1037/hea0000475

Terracciano, A., McCrae, R. R., Brant, L. J., & Costa, P. T., Jr. (2005). Hierarchical linear modeling analyses of the NEO-PI-R scales in the Baltimore Longitudinal Study of Aging. *Psychology and Aging*, *20*(3), 493–506. https://doi.org/10.1037/0882-7974.20.3.493

Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, *90*(2), 288–307. https://doi.org/10.1037/0022-3514.90.2.288

Tsai, J. L., Louie, J. Y., Chen, E. E., & Uchida, Y. (2007). Learning what feelings to desire: Socialization of ideal affect through children's storybooks. *Personality and Social Psychology Bulletin*, *33*(1), 17–30. https://doi.org/10.1177/0146167206292749

Wang, D., & Cui, H. (2003). Processes and preliminary results in the construction of the Chinese personality scale (QZPS). *Acta Psychologica Sinica*, *35*(1), 127–136.

Wang, D., Cui, H., & Zhou, F. (2005). Measuring the personality of Chinese: QZPS versus NEO PI-R. *Asian Journal of Social Psychology*, *8*(1), 97–122. https://doi.org/10.1111/j.1467-839X.2005.00158.x

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of*

Personality and Social Psychology, *37*(3), 395–412. https://doi.org/10.1037/0022-3514.37.3.395

Wood, A. M., Joseph, S., & Maltby, J. (2009). Gratitude predicts psychological well-being above the Big Five facets. *Personality and Individual Differences*, *46*(4), 443–447. https://doi.org/10.1016/j.paid.2008.11.012

Wu, K., Lindsted, K. D., Tsai, S. Y., & Lee, J. W. (2008). Chinese NEO-PI-R in Taiwanese adolescents. *Personality and Individual Differences*, *44*(3), 656–667. https://doi.org/10.1016/j.paid.2007.09.025

Xi, C., Zhong, M., Lei, X., Liu, Y., Ling, Y., Zhu, X., Yao, S., & Yi, J. (2018). Psychometric properties of the Chinese version of the neuroticism subscale of the NEO-PI. *Frontiers in Psychology*, *9*, Article 1454. https://doi.org/10.3389/fpsyg.2018.01454

Xia, L. X., & Huang, X. T. (2008). Development of the Self-supporting Personality Scale for Adolescent Students. *Acta Psychologica Sinica*, *40*(5), 593–603.

Yang, J. F. (2010). Cross-cultural personality assessment: The Revised NEO Personality Inventory in China. *Social Behavior and Personality*, *38*(8), 1097–1104. https://doi.org/10.2224/sbp.2010.38.8.1097

Yang, J. F., McCrae, R. R., Costa, P. T., Jr., Dai, X., Yao, S., Cai, T., & Gao, B. (1999). Cross-cultural personality assessment in psychiatric populations: The NEO-PI-R in the People's Republic of China. *Psychological Assessment*, *11*(3), 359–368. https://doi.org/10.1037/1040-3590.11.3.359

Yik, M. (2009). Studying affect among the Chinese: The circular way. *Journal of Personality Assessment*, *91*(5), 416–428. https://doi.org/10.1080/00223890903087596

Yik, M. (2010a). Affect and interpersonal behaviors: Where do the circumplexes meet? *Journal of Research in Personality*, *44*(6), 721–728. https://doi.org/10.1016/j.jrp.2010.10.003

Yik, M. (2010b). How unique is Chinese emotion? In M. H. Bond (Ed.), *The Oxford handbook of Chinese psychology* (pp. 205–220). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541850.001.0001

Yik, M. (2022). [Unpublished raw data on the relationships between NEO-FFI-3 and happiness]. Hong Kong University of Science and Technology.

Yik, M., & Bond, M. H. (1993). Exploring the dimensions of Chinese person perception with indigenous and imported constructs: Creating a culturally balanced scale. *International Journal of Psychology*, *28*(1), 75–95. https://doi.org/10.1080/00207599308246919

Yik, M., Mues, C., Sze, I. N. L., Kuppens, P., Tuerlinckx, F., De Roover, K., Kwok, F. H. C., Schwartz, S. H., Abu-Hilal, M., Adebayo, D. F., Aguilar, P., Al-Bahrani, M., Anderson, M. H., Andrade, L., Bratko, D., Bushina, E., Choi, J. W., Cieciuch, J., Dru, V., . . . Russell, J. A. (2022). On the relationship between valence and arousal in samples across the globe. *Emotion*. Advance online publication. https://doi.org/10.1037/emo0001095

Yik, M., & Russell, J. A. (2001). Predicting the Big Two of affect from the Big Five of personality. *Journal of Research in Personality*, *35*(3), 247–277. https://doi.org/10.1006/jrpe.2001.2322

Yik, M., & Russell, J. A. (2004). On the relationship between circumplexes: Affect and Wiggins' IAS. *Multivariate Behavioral Research*, *39*(2), 203–230. https://doi.org/10.1207/s15327906mbr3902_4

Yik, M., Russell, J. A., Ahn, C. K., Fernández Dols, J. M., & Suzuki, N. (2002). Relating the Five-Factor Model of personality to a circumplex model of affect: A five language study. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 79–104). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0763-5_5

Yik, M., Russell, J. A., & Steiger, J. H. (2011). A 12-point circumplex structure of core affect. *Emotion*, *11*(4), 705–731. https://doi.org/10.1037/a0023980

Zhang, B., Li, Y. M., Li, J., Luo, J., Ye, Y., Yin, L., Chen, Z., Soto, C. J., & John, O. P. (2022). The Big Five Inventory–2 in China: A comprehensive psychometric evaluation in four diverse samples. *Assessment*, *29*(6), 1262–1284. https://doi.org/10.1177/10731911211008245