PERSONALITY
Science

Empirical

Perceived Weirdness: A Multitrait-Multisource Study of Self and Other Normality Evaluations

Jun-Yeob Kim[1], Daniel A. Newman[1,2], P. D. Harms[3], Dustin Wood[3]

[1] *Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA.* **[2]** *School of Labor & Employment Relations, University of Illinois at Urbana-Champaign, Champaign, IL, USA.* **[3]** *Department of Management, University of Alabama, Tuscaloosa, AL, USA.*

Corresponding Author: Jun-Yeob Kim, 603 East Daniel Street, Champaign, IL, 61820. E-mail: junyeob3@illinois.edu

Abstract

Research in personality and organizational psychology has begun to investigate a novel evaluative trait known as perceived normality, defined as an overall perception that one is normal (vs. strange or weird). The current work evaluates a brief measure of this trait (i.e., a "weirdness scale"), extending past work by assessing both self-reports and peer reports of these normality evaluations. Results confirm the measurement equivalence of self- and peer-reports of perceived weirdness, and discriminant validity of self- and peer-reports of perceived weirdness from Big Five traits. A multitrait-multisource analysis further reveals that trait loadings are larger than self-report and peer-report method loadings for the measure of perceived weirdness. Implications for measurement of self-perceptions and social perceptions of weirdness/normality are discussed.

Keywords

perceived weirdness, normality evaluations, multitrait-multimethod, measurement equivalence, Big Five

> **Relevance Statement**
>
> This paper investigates perceived weirdness from both self and peer perspectives. Perceived weirdness is distinct from Big Five personality, and implications of the trait for future research on norms, culture, and morality are discussed.
>
> **Key Insights**
> - Perceived weirdness is shown to be distinct from Big Five
> - The weirdness scale captures the construct equivalently across self/peer reports.
> - The measure reflects more trait variance than rating-source method variance.

> "So one of our interview questions is, literally, on a scale of 1 to 10, how weird are you?"
>
> Tony Hsieh, Former CEO of *Zappos.com* (New York Times, 2010)

Normality evaluations refer to a distinct dimension of evaluative judgment that entails the overall perception that one is *normal* (vs. s*trange* or *weird*; Benet-Martínez & Waller, 2002; Wood et al., 2007). Although the concept has been studied in the clinical psychology and health literatures (e.g., Genuis & Bronstein, 2017; Offer & Sabshin, 1966; Shoben, 1957; Vaillant, 2003), social, personality, and organizational psychologists have paid less attention to the concept despite its fundamental importance in interpersonal perception (see Wood et al., 2007). To fill this gap, Wood et al. (2007) investigated the nature of normality evaluations by using evaluative lexical terms (i.e., adjectives commonly used for describing people) including normality adjectives (e.g., *normal, weird*), and found that normality adjectives form a separate factor called *perceived normality* or *perceived weirdness* (although we use the terms perceived normality and perceived weirdness interchangeably, we prefer the perceived weirdness label, given the adjective *weird* exhibits a much higher loading on the latent factor, compared to the adjective *normal*). Wood et al. also discussed the relations of perceived weirdness with Big Five personality traits. Although Wood and colleagues' work has contributed key insights to the concept of weirdness evaluations, it has not spurred much follow-up in terms of empirical research.

One possible reason might be that the measurement of perceived weirdness has not been extensively validated. In particular, very little is known about the convergence between self-perceived weirdness and peer perceptions of one's weirdness. Using self-reports of descriptive and normality-related adjectives, Wood et al. (2007) conducted two confirmatory factor analyses (CFA) and found that weirdness/normality is a distinct dimension. However, the five factors identified in their first CFA (i.e., *decent/good, exciting/extraordinary, awful/ridiculous, strange vs. normal*, and *snobbish/aggravating*) were considered apart from the Big Five framework (Extraversion, Agreeableness, Openness to Experience, Emotional Stability, and Conscientiousness; Goldberg, 1992). Since Wood and colleagues only used evaluative adjectives in their CFA, their analyses may have

PsychOpen GOLD

missed the overlaps with other traits that describe individuals; thus, it is unclear to what extent perceived weirdness relates nomologically to Big Five traits (including convergent/discriminant validity). Furthermore, Wood et al.'s arguments (2007) were based on self-assessments of normality evaluation (i.e., how individuals perceive or assess *themselves* as weird/normal), and it as such remains unclear how self-reports of perceived weirdness might differ from non-self-reported, peer perceptions of weirdness.

In the current study, we seek to offer several contributions related to the construct validity of perceived weirdness, both as a self-reported and a peer-reported trait. First, we provide an original confirmation of the factor structure among the six self-reported personality traits (i.e., perceived weirdness and the Big Five personality traits; which are measured using adjectival items, as described in the Method section). In so doing, we are able to investigate the convergent and discriminant validity (i.e., convergent validity between indicators of the same construct, and discriminant validity between latent constructs; Bagozzi & Edwards, 1998) of perceived weirdness, and the nomological validity of perceived weirdness with regard to the Big Five domains. Second, we extend our analysis by confirming the factor structure among perceived weirdness and Big Five traits beyond self-report methods, using peer-report methods. Third, we combine the self-reported and peer-reported data into a single analysis, to establish that self-reported perceived weirdness and peer-reported perceived weirdness can be conceptualized to represent two distinct constructs (e.g., distinguishing between personality identity and personality reputation; Hogan, 1991). Fourth, we attempt to establish measurement equivalence of normality evaluations across self- and peer-perceptions (Vandenberg & Lance, 2000). Finally, we conduct a multitrait-multimethod (MTMM) confirmatory factor analysis to test different models that consist of six trait factors (i.e., perceived weirdness and Big Five personality traits) and two method factors (i.e., self-report and peer-report). Overall, this work attempts to establish construct-valid measurement of the perceived weirdness/normality trait, from self and peer perspectives.

Weirdness/Normality Evaluations

Normality has often been understood in terms of *abnormality* (i.e., not being normal), because of the ease of defining abnormality (compared to defining normality; Wood et al., 2007). Despite several researchers' requests for a clear definition of normality (e.g., Offer & Sabshin, 1966, 1991; Shoben, 1957; Vaillant, 2003), there had been little research on normality evaluations until Wood et al.'s (2007) investigation. These authors performed exploratory principal components analyses to examine whether normality evaluations represent a distinct dimension of evaluative judgment, analyzing the 92 trait adjectives from Saucier's (1997) common trait adjectives that were identified as highly evaluative, alongside both synonyms (e.g., *average*, *normal*, and *ordinary*) and antonyms (e.g., *weird*, *abnormal*, *exceptional*, *extraordinary*, *original*) of the English words *normal* and *average*. The exploratory results suggested that these adjectives separately loaded

onto two factors, which were ultimately labeled *perceived normality* (i.e., weird, strange, normal, abnormal) and *perceived uniqueness* (e.g., extraordinary, remarkable, exceptional, unique). Wood et al. (2007) then dropped perceived uniqueness from further analysis to focus on perceived normality only, likely because of its nomological validity (self-reported perceived normality/weirdness was a correlate of fitting in with peers, whereas perceived uniqueness was not) and its discriminant validity (self-reported perceived normality/weirdness showed adequate discriminant validity from Big Five traits, whereas perceived uniqueness was strongly overlapping with Openness to Experience).

Focusing on perceived normality/weirdness, the authors concluded that being normal captures positive aspects of being "standard or usual." They claimed, "normality evaluations reflect an individual's own determination of whether his or her pattern of behavior is socially acceptable or whether it is unacceptable and should be altered" (Wood et al., 2007, p. 862), noting that norms or normative social forces have been understood as among the reasons for individuals' behavior and psychological development across the lifespan (e.g., Ajzen, 1991; Roberts et al., 2005). Based on this argument, Wood and colleagues (2007) found that individuals who scored low on perceived normality (i.e., people who perceive themselves as more weird/less normal) felt a stronger need or desire to improve their personality, whereas individuals with high normality evaluations (i.e., who perceive themselves as less weird) tended to think they fit better with their peers. Although Wood et al.'s findings contributed to the understanding of perceived weirdness as a trait construct, we note that their arguments and findings are exclusively based on self-perceptions of weirdness. Therefore, the current study seeks to investigate both measurement equivalence and convergence between self-perceptions and peer-perceptions of normality evaluations.

Weirdness/Normality Evaluations and Big Five Personality

To understand the relations between normality evaluations and Big Five personality traits, Wood et al. (2007) drew upon diverse perspectives, echoing Offer and Sabshin (1966, 1991). As such, Wood et al. proposed that normality evaluations would correlate positively with Big Five personality traits that substantially relate to norm-adherence and conventionality (i.e., Agreeableness, Conscientiousness, and low Openness to experience: Benet-Martínez & Waller, 2002; De Raad & Barelds, 2008; Simms, 2007), as well as well-being or mental health (i.e., Emotional Stability [Neuroticism], Conscientiousness, Extraversion, and Agreeableness: see Kotov et al., 2010). In the end, Wood et al. (2007) reported the correlations between self-reported perceived weirdness (i.e., participants rated their own perceived weirdness) and both self- and peer-reported Big Five personality traits (i.e., peers rated the target's personality), and found that self-perceived weirdness was negatively related to Agreeableness, Conscientiousness, and Emotional Stability, positively related to Openness to Experience, and unrelated to Extraversion (results replicated across both self-reported and peer-reported Big Five traits).

PsychOpen GOLD

Interestingly, past research does not appear to have investigated the factor structure among weirdness evaluations and the Big Five personality traits analyzed together, which is an important step for establishing the discriminant validity of perceived weirdness. We thus conducted a series of CFA (Steps 1 and 2) and an MTMM (Step 3) analyses by using both self- and peer-reported data, to reveal the structure among those six personality traits, to provide evidence of convergent and discriminant validity of weirdness/normality evaluations, and to partition variance in these measures into trait and method components. These analyses also allow us to estimate the relationships between perceived weirdness and Big Five traits, when both perceived weirdness and the Big Five traits are measured using both self- and peer-report.

Measurement Equivalence

In addition to using CFA to establish convergent and discriminant validity in the measurement of perceived weirdness, and to partitioning trait variance from method variance in these personality measures, we also seek to assess measurement equivalence between self- and peer-reports of these measures. Vandenberg and Lance (2000) summarized a sequence of steps for establishing measurement equivalence, using structural equation modeling (SEM). The first step, configural invariance, tests whether the groups have the same general factor structure (pattern of factor loadings). This step requires specifying the same factor structure within each condition (self- and peer-report) separately, allowing all model parameters to differ across the two conditions (the model can be evaluated by fit indices such as RMSEA, SRMR, TLI, and CFI). Next, metric invariance should be tested, by constraining the previous model to have equal factor loadings across conditions (self- and peer-report). Afterward, scalar invariance can be tested, by constraining the intercepts for each indicator to be equal across conditions. Therefore, nested models with equal factor structure, equal factor loadings, and equal intercepts across conditions can be compared.

Method

Participants and Procedure

Participants were recruited from seven student organizations at a large Midwestern University (56% female, mean age = 19.54). We asked participants to rate their own traits (including adjectives measuring Big Five personality and weirdness/normality evaluations). Next, each participant was asked to rate three peers from their same organization, using the same adjectives that were used for the self-ratings. We note that the peers were selected randomly, within each organization. Each participant received $10 monetary compensation. Overall, 370 participants provided self-ratings and 436 participants rated their peers. Sample size was predetermined (archival dataset), but was notably larger

than *N* for similar past MTMM CFA analyses (Joseph & Newman, 2010). On average, 2.26 peers (*SD* = .95) rated each participant. Rather than using listwise deletion and dropping partially incomplete cases, all data were included in the analyses using a FIML missing data technique (Newman, 2014). This sample was used in prior studies, but which reported on different combinations of the variables: Harms et al. (2007) used self-rated Big Five, but no normality evaluations nor any peer-rated data; Wortman and Wood (2011) only used self-rated data, and did not report peer-rated normality evaluations nor relationships between normality evaluations and Big Five traits; and Kim et al. (2020) only used peer-rated normality, but not self-rated normality nor Big Five traits. As such, the correlations analyzed in the current paper did not appear in past studies.

Measures

Instructions for the measures were adapted from Goldberg (1992). For self-report, we asked, "How do you see yourself in general? Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself *generally* or *typically*, and as you see yourself at the present time, not as you wish to be in the future." For peer-report, we asked, "How would you describe this person's personality? Describe this person as accurately as possible, as you see him or her at the present time, not as they wish to be in the future. Describe this person as he or she is *generally* or *typically*."

Perceived Weirdness/Normality

Perceived weirdness[1] was measured with a six-item scale as reported in Kim et al. (2020), adapted from Wood et al. (2007). Using this 'weirdness scale,' we asked participants to rate themselves, and they also received peer ratings, on perceived weirdness. Participants read the sentences, "I see myself as…", or "I see this person as…", followed by the trait adjectives: weird, normal, abnormal, odd, strange, and unusual ('normal' was reverse coded[2]). Each adjective was rated on a 5-point scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*; self: α = .84; peer: α = .87).

---

1) A helpful reviewer questioned how our weirdness measure might relate to measures of abnormal personality dimensions. Unfortunately, we do not have any published measures of abnormal personality dimensions in our data, so this question is left for future research.

2) A helpful reviewer noted that only one item in the perceived weirdness scale (i.e., the item normal) was reverse-worded. Although reverse-worded items may reduce reliability and validity of a scale under certain conditions (Schmitt & Stults, 1985; Woods, 2006), our current data showed that this item had acceptable factor loadings and the reliability of perceived weirdness was acceptable (across self- and peer-reports). We also conducted the same analyses including fretful. Although the reliability of peer-rated emotional stability dropped to .62, all other analyses suggested the same results. Detailed results can be provided upon request.

Big Five Personality

We measured the Big Five traits using adjectives from Goldberg (1992). For four Big Five traits (i.e., Extraversion/Surgency, Agreeableness, Conscientiousness, and Openness/Intellect) we used ten trait adjectives per Big Five domain, which Goldberg sampled to include the first five adjectives from both the positive and negative poles of each trait (Goldberg, 1992, pp. 34–35). For Emotional Stability/Neuroticism, however, Goldberg reported several adjectives that either had low factor loadings or that cross-loaded onto another factor in his self-report or peer-report data—we chose to avoid these items and instead used the first seven items reported by Goldberg that had average loadings above .4 on the intended factor and that did not load onto another factor above .3, as reported in factor analyses in Goldberg's original study. Furthermore, we found that the adjective 'fretful' showed a negative item-total correlation in the current sample, and we thus decided to exclude fretful from our analyses[3]. These six remaining adjectives were: relaxed, unenvious, anxious, moody, envious, and jealous (all items are listed with the factor analysis results below in Table 2).

Participants rated themselves and their peer(s) on the Big Five adjectives. Participants read the sentences, "I see myself as…", or "I see this person as…" and the Big Five personality adjectives followed. Each adjective was rated on a 5-point scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*). Reliabilities were: Extraversion (self: α = .87; peer: α = .90), Agreeableness (self: α = .79; peer: α = .92), Conscientiousness (self: α = .82; peer: α = .87), Emotional Stability (self: α = .72; peer: α = .70), and Openness to Experience (self: α = .78; peer: α = .81).

Transparency, Openness, and Reproducibility

The current study is not pre-registered. The hypothesized models were all *a priori* and confirmatory (there were no *post hoc* model modifications, other than to the Emotional Stability measure–see Measures section above). R code and raw data for reproducing the results are available online (see Supplementary Materials). As for additional analyses that are not reported, we did attempt to estimate models with item-level indicators for the Big Five, but those models did not converge; so we then used parcels (i.e., means across items) as indicators, where items were assigned to parcels only once, using a random number generator in R. There were no data exclusions, and no alternative measures of the studied variables were analyzed.

---

3) We also conducted the same analyses including fretful. Although the reliability of peer-rated emotional stability dropped to .62, all other analyses suggested the same results. Detailed results can be provided upon request.

PsychOpen GOLD

Analyses and Results

Table 1 shows descriptive statistics and bivariate correlations. Using the sample and measures described above, we first conducted separate self-reported CFA and peer-reported CFA, then combined them for the analyses of self- and peer-reported CFA together, followed by the assessment of measurement equivalence across sources (Weirdness-Big Five CFA and measurement equivalence). Next, we conducted MTMM analyses by specifying a single CFA model with trait and method factors (Widaman, 1985), then partitioned trait and method variance in each trait-method unit.

Table 1

Correlations Among Self- and Peer-Reported Perceived Weirdness and Big Five Personality

Variable	1	2	3	4	5	6	7	8	9	10	11	12
Self-report												
1. Perceived Weirdness	.84											
2. Openness to Experience	.11	.78										
3. Agreeableness	-.18	.32	.79									
4. Conscientiousness	-.24	.14	.33	.82								
5. Extraversion	-.15	.40	.23	.08	.87							
6. Emotional Stability	-.15	.12	.24	.12	.11	.72						
Peer-report												
7. Perceived Weirdness	.24	.10	-.17	-.09	.01	.11	.87					
8. Openness to Experience	.01	.17	.08	.14	.18	-.07	-.32	.81				
9. Agreeableness	-.00	.00	.21	.06	.03	-.02	-.47	.59	.92			
10. Conscientiousness	-.13	.00	.09	.30	.03	.00	-.53	.57	.62	.87		
11. Extraversion	-.02	.15	.07	-.02	.51	.02	-.00	.35	.07	-.05	.90	
12. Emotional Stability	-.04	-.04	.14	.02	-.09	.06	-.37	.27	.58	.34	-.11	.70

Note. N = 339–436. Reliability in the diagonal; correlations $|r| \geq .11$ are statistically significant ($p < .05$). We note that the observed self-other correlations for Big Five traits reported in Table 1 generally ranged in magnitude from .2 to .5, which is in line with past meta-analytic evidence for the magnitude of self-other Big Five correlations among cohabitors (Connelly & Ones, 2010)—with a single exception. Our current self-other correlation for Emotional Stability was remarkably small ($r = .06$). We thus urge due caution in interpreting the generalizability of our results involving Emotional Stability.

Step 1: CFA of Self-Reported and Peer-Reported Data

Analysis

We first conducted CFA on perceived weirdness and the Big Five trait measures. Three *a priori* CFA specifications were estimated: (a) Self-report CFA: using only the self-reported data (oblique 6-factor model), (b) Peer-report CFA: using only the peer-reported data (oblique 6-factor model), and (c) Self- and peer-report CFA: using the self- and peer-reported combined data (oblique 12-factor model). Results of these analyses appear in

Table 2[4]. The perceived weirdness measure was factor analyzed with items as indicators, whereas each of the Big Five traits was analyzed by assembling its items into three parcels. Parceling has the advantage of creating indicators that are more reliable, more normally distributed, and more granular, while tremendously reducing the number of parameters that must be estimated (Williams et al., 2009). Nonetheless, we do not parcel the perceived weirdness items because we are still interested in item-level diagnostic information on the weirdness measure. Items were assigned to parcels randomly using R Studio (see parcels in Table 2).

Table 2

Step 1: CFA of Perceived Weirdness and Big Five Personality

	Factor Loadings		
Observed Variables	**Self-Report** *(Oblique 6-factor)*	**Peer-Report** *(Oblique 6-factor)*	**Combined Self/Peer** *(Oblique 12-factor)*
Perceived Weirdness/Normality (Wood et al., 2007)			
Weird	.79	.78	.78/.78
Strange	.79	.80	.79/.80
Odd	.75	.78	.75/.78
Abnormal	.69	.78	.69/.78
Normal	-.42	-.50	-.42/-.50
Unusual	.70	.72	.71/.72
Big Five Personality (Goldberg, 1992)			
P1 (Extrav: Verbal, Shy, Quiet)	.78	.89	.79/.88
P2 (Extrav: Extraverted, Reserved, Energetic)	.83	.85	.83/.85
P3 (Extrav: Assertive, Introverted, Talkative, Untalkative)	.86	.88	.87/.89
P4 (Agree: Unsympathetic, Harsh, Cooperative)	.72	.89	.73/.89
P5 (Agree: Distrustful, Cold, Warm)	.71	.89	.70/.89

---

4) A helpful reviewer asked whether the distinction between normal vs. weird might be confounded with general positive vs. negative evaluative meaning. In response to this concern, we conducted a supplemental analysis in which we correlated peer ratings of perceived weirdness with peer ratings of liking (using one available item in our dataset, "Rate the extent to which you like each member of the organization," from 1 = strongly dislike to 7 = strongly like). The correlation between peer-rated weirdness and liking was only $r = -.16$ (and the corresponding correlation between self-rated weirdness and peer-rated liking was only $r = -.03$), suggesting that although weirdness may contain some social desirability, it is far from redundant with the general factor of positive vs. negative evaluation/liking.

	Factor Loadings		
Observed Variables	Self-Report (*Oblique 6-factor*)	Peer-Report (*Oblique 6-factor*)	Combined Self/Peer (*Oblique 12-factor*)
P6 (Agree: Kind, Sympathetic, Trustful, Unkind)	.77	.90	.77/.90
P7 (Cons: Disorganized, Careless, Unsystematic)	.84	.81	.84/.82
P8 (Cons: Inefficient, Organized, Thorough)	.85	.85	.86/.85
P9 (Cons: Neat, Practical, Systematic, Undependable)	.83	.91	.82/91
P10 (Open: Creative, Unintellectual, Unintelligent)	.81	.89	.81/.88
P11 (Open: Uncreative, Bright, Imaginative)	.81	.84	.82/.85
P12 (Open: Complex, Intellectual, Simple, Unimaginative)	.57	.54	.57/.54
P13 (Emot. Stab.: Anxious, Fretful, Envious)	.90	.60	.87/.59
P14 (Emot. Stab.: Moody, Relaxed)	.46	.80	.49/.81
P15 (Emot. Stab.: Jealous, Unenvious)	.62	.51	.64/.50
Fit Indices			
χ^2 (*df*)	407.66 (174)	639.69 (174)	1503.70 (753)
RMSEA/SRMR	.060/.056	.078/.062	.046/.055
TLI/CFI	.91/.93	.90/.92	.91/.92

Note. N = 367 ratees (self-report), N = 436 ratees (peer-report). Missing data treatment = Full Information Maximum Likelihood (FIML). P = Parcel. For combined data, loadings before the slash (/) are self-report items loaded onto self-report traits, after the slash (/) are peer-report items loaded onto peer-report traits.

Results

Results of Self-report, Peer-report, and Self-and-Peer-report CFA models showed similar fit indices, factor loadings, and factor intercorrelations (see Table 2). All three CFA models produced model fit indices that we deemed acceptable. In addition, all standardized factor loadings were larger than .41. The average factor correlation was ϕ = .24 for the self-report data, ϕ = .39 for the peer-report data, and ϕ = .18 for the combined data (see observed correlations in Table 1). Together, these results confirm the oblique solution among perceived weirdness and the Big Five traits, and support perceived weirdness as a distinct construct from the Big Five traits.

Step 2. Measurement Equivalence

Analysis

We next attempted to establish measurement equivalence across the two measurement sources (i.e., self- and peer-reports) following Vandenberg and Lance's (2000) guidelines. Three types of oblique 12-factor models were compared: (a) Model 1 (configural invariance) establishes same pattern of factor loadings (whether an item loads vs. does not load onto each factor) across self- and peer-reports, (b) Model 2 (metric invariance) constrains factor loadings to be equal across self- and peer-reports, and (c) Model 3 (scalar invariance) constrains item intercepts to be equal across self- and peer-reports. For Model 3, we specified three sequential models: Model 3a (partial scalar invariance, constraining equal intercepts for perceived weirdness, but allowing unequal intercepts for the Big Five indicators), Model 3a' (partial scalar invariance, with equal intercepts for Big Five indicators, but allowing unequal intercepts for perceived weirdness), and Model 3b (scalar invariance, with equal intercepts for all indicators of both Big Five and perceived weirdness). The sequence of models (Models 1–3) is nested, with each model progressively imposing additional constraints. We compared model fit indices and used sequential ΔCFI tests (Cheung & Rensvold, 2002) to establish measurement equivalence. For identification, we fixed one indicator loading to 1.0 for each latent factor.

Results

Table 3 shows the fit indices of Models 1–3. Regarding absolute model fit, we judge Model 1 (configural invariance), Model 2 (metric invariance), and Model 3a (partial scalar invariance for perceived weirdness) to exhibit adequate fit, while Model 3a' (partial scalar invariance for Big Five) and Model 3b (scalar invariance) exhibit sub-optimal absolute fit. Namely, these results support metric equivalence/equal factor loadings between self- and peer-report, for both perceived weirdness and the Big Five personality traits (Model 2), as well as partial scalar invariance/equal intercepts between self- and peer-report, for perceived weirdness (Model 3a). Further, partial scalar invariance (equal intercepts for the Big Five) does not appear to be supported for the Big Five (Model 3a'; ΔCFI = .018; see Table 3) in the current work. In sum, this study provides initial evidence for both metric and scalar equivalence of the perceived weirdness measure across self- and peer-report.

Table 3

Step 2: Measurement Equivalence Between Self- and Peer-Report (Oblique 12-Factor Model)

Measurement Equivalence Model	χ^2	df	RMSEA	TLI	CFI	SRMR	ΔCFI
Model 1. Configural Invariance	1503.70	753	.046	.907	.919	.055	—
Model 2. Metric Invariance (Equal Loadings)	1596.97	768	.048	.900	.911	.059	.008
Model 3a. Partial Scalar Invariance (Equal Intercepts for Weirdness, Unequal intercepts for Big Five)	1636.05	773	.049	.896	.907	.060	.004
Model 3a'. Partial Scalar Invariance (Equal Intercepts for Big Five, Unequal intercepts for Weirdness)	1773.67	778	.053	.881	.893	.062	.018
Model 3b. Scalar Invariance (Equal Intercepts)	1812.77	783	.053	.878	.889	.063	.022

Note. N = 464 ratees. Missing data treatment = Full Information Maximum Likelihood (FIML). All Δ χ^2 are statistically significant (*p* < .05).

Step 3. Multitrait-Multimethod Analysis

Analysis

Next, we conducted an MTMM analysis in the CFA framework to assess discriminant and convergent validity of the perceived weirdness measure (Widaman, 1985). In our MTMM analyses, we specified six trait factors (for perceived weirdness and Big Five personality traits), and two method factors (for both self-report [participants rated their own personality traits] and peer-report [peers rated the targets' personality traits]). In particular, we specified that each trait-method latent factor from the 12-factor model would double load, onto two corresponding second-order traits—a trait factor and a method factor (see Figure 1). For example, the self-report extraversion factor was specified to double-load, once onto the extraversion trait factor, and once onto the self-report method factor. To achieve model identification, we constrained the two trait loadings for each trait to equality (e.g., extraversion [self] and extraversion [peer] loadings onto the extraversion trait factor were set equal), and we also fixed the factor loadings to 1.0 for agreeableness [self] onto the self-report method factor, and for agreeableness [peer] onto the peer-report method factor.

We compared seven models (see Widaman, 1985) that are combinations of different trait factor structures (i.e., no trait factor, correlated trait factors, or one general trait factor) and method factor structures (i.e., no method factor, self- and peer- report meth-

od factors uncorrelated, or method factors correlated; see Table 4). Trait factors were allowed to intercorrelate, but trait and method factors were constrained to be uncorrelated. Convergent validity (i.e., extent to which the scales designed to assess the same construct are strongly related) can be established via large trait loadings in the MTMM model. For instance, if self- and peer-reports of perceived weirdness exhibit large average trait loadings onto the weirdness trait, this is consistent with convergent validity. Discriminant validity (i.e., extent to which scales designed to assess different constructs are not too strongly related) can be demonstrated by assessing the correlations among latent traits in the MTMM model. For instance, if perceived weirdness and Big Five traits are correlated notably less than unity, it suggests discriminant validity. We estimated a subset of Widaman's (1985), as implemented by Joseph and Newman (2010), models to demonstrate convergent and discriminant validity.

Table 4

Step 3: MTMM Results for Perceived Weirdness and Big Five Personality

| MTMM Model	Trait Factors	Method Factors	χ^2	df	RMSEA	TLI	CFI	SRMR
Model I	Correlated	None	2067.46	798	.059	.85	.86	.120
Model II	Correlated	Uncorrelated	1623.35	786	.048	.90	.91	.065
Model III	Correlated	Correlated	1618.40	785	.048	.90	.91	.064
Model IV	None	Uncorrelated	5994.37	819	.117	.41	.44	.133
Model V	None	Correlated	5988.86	818	.117	.41	.44	.130
Model VI	1 General trait	None	6885.82	819	.126	.31	.35	.148
Model VII (Weird.-Consc. -1.0)	Correlated	Correlated	2155.17	786	.061	.84	.85	.870

Note. N = 464 ratees. MTMM Models are estimated in hierarchical CFA models, with 12 first-order factors, and trait and method higher-order factors specified to model relationships among the 12 first-order factors. Missing data treatment = Full Information Maximum Likelihood (FIML). Model VII constrained the correlation between perceived weirdness and conscientiousness to be 1.0. Best fitting model (Model III) is in italics.

Figure 1 depicts Model III. All indicators loaded onto their corresponding first-order trait-method latent factors (i.e., 12 factors: normality plus Big Five × self- and peer-report). Then, these 12 latent factors loaded onto both trait and method higher-order factors. The left side of the figure represents trait factors: each trait-method latent factor loaded on its corresponding trait factor (e.g., both self- and peer-reported perceived weirdness loaded onto the perceived weirdness trait factor). On the right side of Figure 1 are the method factors (e.g., all self-report trait-method latent factors loaded onto the Self-Report method factor; see Figure 1).

Figure 1

Step 3: MTMM Model III for Normality Evaluations and Big Five Personality
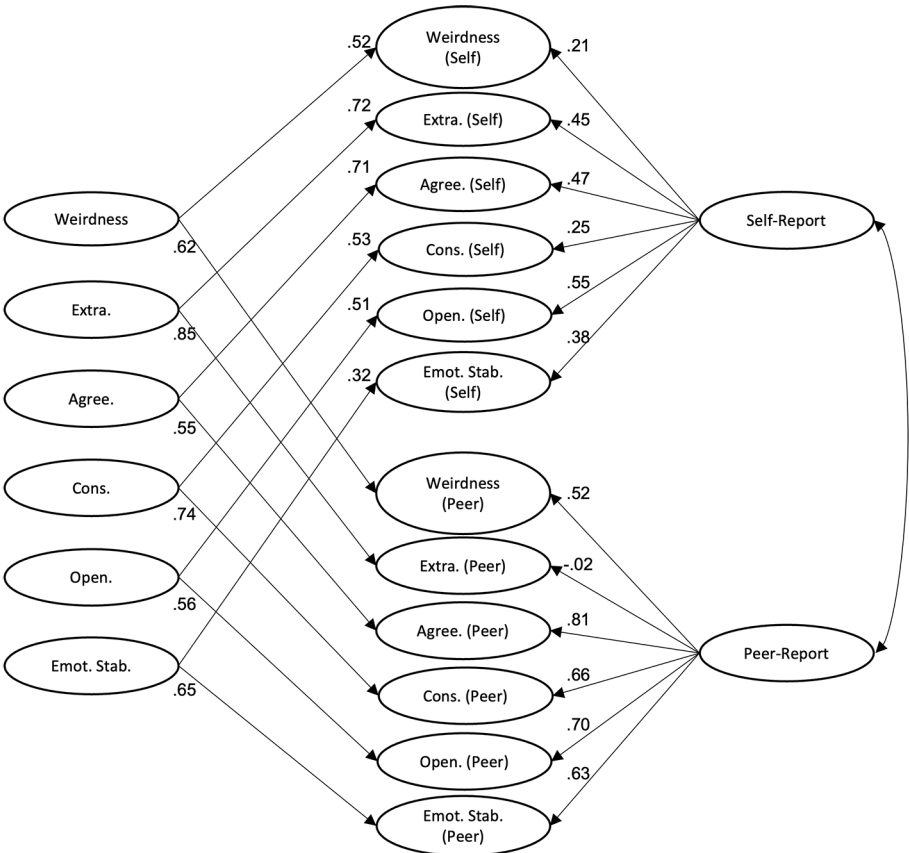
Note. Extra. = Extraversion; Agree. = Agreeableness; Cons. = Conscientiousness; Emot. Stab. = Emotional Stability; Open. = Openness to Experience; Self = Self-Report; Peer = Peer-Report. All second-order trait factors (factors on the left side) were correlated although not depicted. Factor loadings and correlations are listed in Table 5a, 5b, and 5c. Indicators and their paths are omitted in this Figure for brevity.

Results

Results of MTMM analyses appear in Table 4. In terms of absolute model-data fit, only Model II (correlated traits, uncorrelated methods) and Model III (correlated traits, correlated methods) exhibited adequate fit, and they also exhibited nearly identical fit indices. In terms of relative fit, these two models both fit notably better than alternative models with no method factors (Model I: ΔCFI = .05; and Model VI: ΔCFI = .56), and alternative models with no trait factors (Model IV: ΔCFI = .47; and Model V: ΔCFI = .47).

These relative fit comparisons confirm that the data are consistent with the existence of both trait factors and method factors.

Next, because Model II (correlated traits, uncorrelated methods) and Model III (correlated traits, correlated methods) both showed adequate and nearly-equivalent fit, we decided to base our interpretations on Model III, because the method correlation (ϕ = -.37) was statistically significant. For Model III (i.e., six oblique trait factors for Big Five personality traits and perceived weirdness, plus two correlated method factors for self-report and peer-report) parameter estimates are shown in Table 5a, 5b, and 5c. As seen in Table 5b, all twelve trait-method factors had substantial trait loadings (> .50) onto their corresponding higher-order trait factors, with the single exception of self-reported Emotional Stability, which loaded at .32 onto its higher-order trait factor. Next, as also seen in Table 5b, the six trait-method factors that were self-reported all had loadings onto their higher-order method factor (i.e., self-report method factor) below .50, with the single exception of self-reported Openness, which loaded at .54 onto the self-report method factor. The average % method variance in the self-report factors was 16%, and the self-report perceived weirdness measure exhibited only 4% method variance (Table 5b). In contrast, the six trait-method factors that were peer-reported all had loadings onto their higher-order method factor (peer-report method factor) above .50, with the single exception of peer-reported Extraversion, which had zero loading onto the peer-report method factor. The average % method variance in the peer-report factors was 36%, and the peer-report perceived weirdness measure exhibited 27% method variance. To summarize the Step 2 MTMM results: (a) the trait loadings were generally large, (b) the self-report method loadings were generally smaller than their corresponding trait loadings, (c) the peer-report method loadings were generally similar in magnitude to their corresponding trait loadings, and (d) for the perceived weirdness measure, trait loadings were notably larger than method loadings. These results reconfirm the convergent and discriminant validity of perceived weirdness.

Table 5a

Step 3: CFA Results for MTMM Model III – Item Level Factor Loadings

Indicator	Weird. (Self/Peer)	Extrav. (Self/Peer)	Agree. (Self/Peer)	Consc. (Self/Peer)	Open. (Self/Peer)	Emot. Stab. (Self/Peer)
Weird	.78/.78					
Strange	.79/.80					
Odd	.75/.78					
Abnormal	.69/.78					
Normal	-.42/-.51					
Unusual	.70/.72					
Extrav. P1		.82/.87				
Extrav. P2		.84/.84				

Indicator	Weird. (Self/Peer)	Extrav. (Self/Peer)	Agree. (Self/Peer)	Consc. (Self/Peer)	Open. (Self/Peer)	Emot. Stab. (Self/Peer)
Extrav. P3		.87/.89				
Agree. P1			.74/.89			
Agree. P2			.71/.89			
Agree. P3			.75/.90			
Consc. P1				.83/.82		
Consc. P2				.86/.85		
Consc. P3				.82/.91		
Open. P10					.84/.88	
Open. P11					.80/.84	
Open. P12					.57/.54	
Emot. Stab. P13						.84/.58
Emot. Stab. P14						.50/.80
Emot. Stab. P15						.67/.51

Note. N = 464 ratees. P = Parcel; self/peer = factor loadings before [after] slash represent loadings onto corresponding lower-order self-report [peer-report] factors (e.g., .78 in the first column represents the loading of the indicator 'Weird' onto the factor 'Self-Report Weirdness').

Table 5b

Step 3: CFA Results for MTMM Model III – Trait Level Factor Loadings

Latent Factor	Weird.	Extrav.	Agree.	Consc.	Open.	Emot. Stab.	Method Loadings	Method Var %[a]
Self								
Weird.	.52						.21	4
Extrav.		.72					.45	19
Agree.			.71				.47	20
Consc.				.53			.25	6
Open.					.51		.55	30
Emot. Stab.						.32	.38	15
Peer								
Weird.	.62						.52	27
Extrav.		.85					-.02	0
Agree.			.55				.81	66
Consc.				.74			.66	43
Open.					.56		.70	49
Emot. Stab.						.65	.63	33

Note. N = 464 ratees.
[a]Method Var % $\frac{\lambda^2_{method} \times var(Y_{method})}{var(Y)}$.

PsychOpen GOLD

Table 5c

Step 3: CFA Results for MTMM Model III – Latent Factor Correlations

Latent Factor	Weird.	Extrav.	Agree.	Consc.	Open.	Emot. Stab.	Self	Peer
Weirdness	(.22)[a]							
Extrav.	-.07	(.52)[a]						
Agree.	-.40	.13	(.30)[a]					
Consc.	-.60	-.09	.38	(.34)[a]				
Open.	-.15	.52	.57	.47	(.25)[a]			
Emot. Stab.	-.36	-.28	.64	.12	.10	(.22)[a]		
Self	—	—	—	—	—	—	(.19)[a]	
Peer	—	—	—	—	—	—	-.37	(.44)[a]

Note. N = 464 ratees.
[a]Unstandardized factor standard deviations are in the diagonal.

Finally, we note that the latent correlation between perceived weirdness and conscientiousness in Model III was -.60, which affects the discriminant validity inferences regarding perceived weirdness. Thus, we tested a model that constrained the latent correlation between weirdness and conscientiousness to -1.0 (Model VII; see Widaman, 1985). As shown in Table 4, the model fit of Model VII is significantly worse than Model III and therefore provides evidence for discriminant validity of perceived weirdness. To provide an additional test of discriminant validity, we also implemented Fornell and Larcker's (1981) test, which requires that the latent correlation between two factors must be smaller than the square root of the average indicator variance explained by each latent factor (also see Joseph & Newman, 2010). The square root of average variance extracted was .83 for perceived weirdness, and was .76 for Conscientiousness, which are both larger than the latent correlation between weirdness and Conscientiousness of -.60. Thus, discriminant validity is supported, according to both tests.

Discussion

The current research made several contributions to understanding the construct validity of self- and peer-reported perceived weirdness. In Step 1, we first conducted CFA using self- and peer-report data to confirm 12 oblique trait-method factors (i.e., 6 traits: Big Five plus perceived weirdness × 2 methods: self- and peer-report). In Step 2, we established measurement equivalence (both metric equivalence [equal factor loadings] and scalar equivalence [equal item difficulties/intercepts]) between self- and peer-report measures of perceived weirdness, suggesting that the perceived weirdness items assess the weirdness construct in a psychometrically equivalent manner across self- and peer-reports.

Beyond Wood et al.'s (2007) work that emphasized self-reported weirdness evaluations, our current results suggest the measurement validity of using peer-reported perceived weirdness (capturing the same construct in the same manner: equal factor structures, factor loadings, and factor intercepts). That is, self- and peer-reports of normality evaluations are calibrated equivalently and can be inferred to have commensurate meaning across measurement sources (Vandenberg & Lance, 2000). In Step 3, we used MTMM analysis in the CFA framework to confirm the convergent and discriminant validity of perceived weirdness (Widaman, 1985). We confirmed six distinct, oblique traits (i.e., perceived weirdness and the Big Five traits) and two correlated method factors (i.e., self-report and peer-report methods). This supports the inferences that perceived weirdness can be distinguished from the Big Five personality traits and measured with both self- and peer-report.

As mentioned previously, the current research found that perceived weirdness is a distinct dimension of personality from the Big Five traits. This finding enables future research into the social and behavioral outcomes that might be uniquely predicted by weirdness perceptions. For example, we speculate that weirdness could possibly associate with one's creativity (Shalley et al., 2004), business entrepreneurship, or adherence to subjective norms (Ajzen, 1991). Further, our establishment of measurement equivalence highlights the enormous potential of investigating weirdness/normality evaluations from other's perceptions. As recommended by Kim et al. (2020), self- and peer-perceptions of weirdness/normality could be investigated in future research as mechanisms for other norm-violation phenomena, such as moral and ethical violations, or cultural effects (Gelfand et al., 2017). Further, it would be worth investigating whether the current study's findings extend to different cultures. For example, the same behaviors or traits might be perceived as weird in one country/culture but not in others. Beyond assessing the universality of perceived weirdness/normality in other countries, future research might also assess whether this personality trait appears in the lexical structure of languages other than English (McCrae et al., 2002). Furthemore, a helpful reviewer suggested that perceived weirdness/normality evaluations would potentially be related to the Honesty-Humility trait of the HEXACO (Ashton & Lee, 2007), which taps into adherence to moral norms.

In addition, a helpful reviewer asked us to attempt to specify whether weirdness might be a meta-trait (like Digman's, 1997, alpha and beta), an interstitial trait (like altruism in the HEXACO model), or an independent trait (like Honesty in the HEXACO model; Ashton & Lee, 2007). At present, we surmise that weirdness is likely either a meta-trait or an independent/distinct trait, but is not likely an interstitial trait. With respect to its status as a distinct trait, we note that weirdness exhibited adequate discriminant validity from the Big Five, in both self- and peer-reported CFA results, as well as MTMM results. It is also noteworthy that the MTMM results show weirdness correlates most strongly with Agreeableness (φ = -.40), Conscientiousness (φ = -.60),

and Emotional Stability (φ = -.37), which are the three lower-order factors of Digman's higher-order alpha factor (cf. DeYoung et al., 2002, labeled this factor Stability, suggesting it entailed stability in social relationships, motivated behavior, and mood). It is possible that weirdness could be conceptualized as a trait akin to this higher-order factor, but coded in the negative direction (i.e., weirdness is empirically related to Disagreeableness, low Conscientiousness, and Neuroticism). Finally, we posit that weirdness is not an interstitial trait. Inspection of post hoc modification indices for our self-rated CFA and for our peer-rated CFA models both showed that: (a) weirdness items would not notably improve model fit if they were allowed to cross-load onto Big Five factors, and (b) if weirdness items were allowed to cross-load onto the Big Five, none of these items from either self- or peer-CFA models would have exhibited standardized cross-loadings greater than .16 in absolute value. In sum, weirdness does not appear to be an interstitial trait of the Big Five, but it could be conceptualized as either a distinct trait or a meta-trait.

The current study also has several limitations. First, our participants are all college students from a single university, suggesting additional work is needed on the generalizability of current findings. Second, beyond adjective-based Big Five measurement (Goldberg, 1992), it would be helpful to replicate findings with other Big Five measures using statements and behaviors (e.g., Extraversion: "feel comfortable around people"; Agreeableness: "sympathize with others' feelings"; Goldberg et al., 2006). Third, additional work should investigate the mechanisms of person perception that come into play when comparing self- vs. other-perceptions of personality (Vazire & Carlson, 2011).

Conclusions

Our research provided evidence of the measurement equivalence (between self- and peer-rating) of perceived weirdness, convergent validity between self- and peer-ratings of weirdness, and discriminant validity of these evaluations from the Big Five traits. Peer-reported weirdness evaluations capture a similar construct to self-report evaluations of weirdness. Research on peer perceptions of weirdness could potentially supply helpful information about various psychological phenomena related to norm violation (e.g., gender norms, cultural norms, moral norms); however little research has investigated this construct. We hope that our validity evidence for self- and peer-reports of perceived weirdness spurs future work on this fundamental evaluative construct.

PsychOpen GOLD

Supplementary Materials

Data and R codes to reproduce the results are provided as Supplementary Materials (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Kim, J., Newman, D., Harms, P., & Wood, D. (2021a). *Supplementary materials to "Perceived weirdness: A multitrait-multisource study of self and other normality evaluations"* [Dataset, codebook]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5085

Kim, J., Newman, D., Harms, P., & Wood, D. (2021b). *Supplementary materials to "Perceived weirdness: A multitrait-multisource study of self and other normality evaluations"* [Scripts]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5086

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*(1), 45–87. https://doi.org/10.1177/109442819800100104

Benet-Martínez, V., & Waller, N. G. (2002). From adorable to worthless: Implicit and self-report structure of highly evaluative personality descriptors. *European Journal of Personality, 16*(1), 1–41. https://doi.org/10.1002/per.431

Bryant, A. (2010, January 9). *On a scale of 1 to 10, how weird are you?* The New York Times. https://www.nytimes.com/2010/01/10/business/10corner.html

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

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092–1122. https://doi.org/10.1037/a0021212

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33*(4), 533–552. https://doi.org/10.1016/S0191-8869(01)00171-4

De Raad, B., & Barelds, D. P. (2008). A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. *Journal of Personality and Social Psychology, 94*(2), 347–364. https://doi.org/10.1037/0022-3514.94.2.347

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*(6), 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *JMR, Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Gelfand, M. J., Harrington, J. R., & Jackson, J. C. (2017). The strength of social norms across human groups. *Perspectives on Psychological Science, 12*(5), 800–809. https://doi.org/10.1177/1745691617708631

Genuis, S. K., & Bronstein, J. (2017). Looking for "normal": Sense making in the context of health disruption. *Journal of the Association for Information Science and Technology, 68*(3), 750–761. https://doi.org/10.1002/asi.23715

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

Goldberg, L. R., 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

Harms, P. D., Roberts, B. W., & Wood, D. (2007). Who shall lead? An integrative personality approach to the study of the antecedents of status in informal social organizations. *Journal of Research in Personality, 41*(3), 689–699. https://doi.org/10.1016/j.jrp.2006.08.001

Hogan, R. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology: Vol. 2* (pp. 873–919). Consulting Psychologists Press.

Joseph, D. L., & Newman, D. A. (2010). Discriminant validity of self-reported emotional intelligence: A multitrait-multisource study. *Educational and Psychological Measurement, 70*(4), 672–694. https://doi.org/10.1177/0013164409355700

Kim, J.-Y., Hsu, N., Newman, D. A., Harms, P. D., & Wood, D. (2020). Leadership perceptions, gender, and dominant personality: The role of normality evaluations. *Journal of Research in Personality, 87*, Article 103984. https://doi.org/10.1016/j.jrp.2020.103984

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

McCrae, R. R., Costa, P. T., Jr., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional and cross-cultural analyses. *Journal of Personality and Social Psychology, 83*(6), 1456–1468. https://doi.org/10.1037/0022-3514.83.6.1456

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods, 17*(4), 372–411. https://doi.org/10.1177/1094428114548590

Offer, D., & Sabshin, M. (1991). Normatology: The next step. In D. Offer & M. Sabshin (Eds.), *The diversity of normal behavior* (pp. 405–419). Basic Books.

Offer, D., & Sabshin, M. (1966). *Normality: Theoretical and clinical concepts of mental health.* Basic Books.

Roberts, B. W., Wood, D., & Smith, J. L. (2005). Evaluating five factor theory and social investment perspectives on personality trait development. *Journal of Research in Personality, 39*(1), 166–184. https://doi.org/10.1016/j.jrp.2004.08.002

Saucier, G. (1997). Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology, 73*(6), 1296–1312. https://doi.org/10.1037/0022-3514.73.6.1296

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9*(4), 367–373. https://doi.org/10.1177/014662168500900405

Shalley, C. E., Zhou, J., & Oldham, G. R. (2004). The effects of personal and contextual characteristics on creativity: Where should we go from here? *Journal of Management, 30*(6), 933–958. https://doi.org/10.1016/j.jm.2004.06.007

Shoben, E. J. (1957). Toward a concept of the normal personality. *The American Psychologist, 12*(4), 183–189. https://doi.org/10.1037/h0045253

Simms, L. J. (2007). The Big Seven model of personality and its relevance to personality pathology. *Journal of Personality, 75*(1), 65–94. https://doi.org/10.1111/j.1467-6494.2006.00433.x

Vaillant, G. E. (2003). Mental health. *The American Journal of Psychiatry, 160*, 1373–1384. https://doi.org/10.1176/appi.ajp.160.8.1373

PsychOpen GOLD

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science, 20*(2), 104–108. https://doi.org/10.1177/0963721411402478

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*(1), 1–26. https://doi.org/10.1177/014662168500900101

Williams, L. J., Vandenberg, R. J., & Edwards, J. R. (2009). 12 structural equation modeling in management research: A guide for improved analysis. *The Academy of Management Annals, 3*(1), 543–604. https://doi.org/10.5465/19416520903065683

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28*(3), 186–191. https://doi.org/10.1007/s10862-005-9004-7

Wood, D., Gosling, S. D., & Potter, J. (2007). Normality evaluations and their relation to personality traits and well-being. *Journal of Personality and Social Psychology, 93*(5), 861–879. https://doi.org/10.1037/0022-3514.93.5.861

Wortman, J., & Wood, D. (2011). The personality traits of liked people. *Journal of Research in Personality, 45*(6), 519–528. https://doi.org/10.1016/j.jrp.2011.06.006