

Ask Once, May Tell: Comparative Validity of Single and Multiple Item Measurement of the Big-Five Personality Factors

**Larry C. Bernard*, R. Patricia Walsh,
& Michael Mills**

Loyola Marymount University

ABSTRACT - Compared validities of single item self-ratings and multiple item questionnaire scores measuring the Big-Five personality domains. Scores on the five domain scales of the *NEO Personality Inventory-Revised* (NEO PI-R; Costa & McCrae, 1992) – Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness – were obtained from 149 participants who also provided self-ratings on descriptions of the same domains. Acquaintances provided criterion ratings of the participants on the NEO PI-R Form R. Multiple regression analyses indicated that participants' domain self-ratings and NEO PI-R scores accounted for equivalent amounts of variance in criterion ratings of Neuroticism and Conscientiousness. However, participants' NEO PI-R scores accounted for significantly more variance in criterion ratings of Extraversion, Openness, and Agreeableness scores than single-item self ratings. Implications are discussed.

When seeking to determine an individual's standing on a personality trait, one commonsensical approach would be to "just ask." Gordon Allport (1941) first raised this possibility when he declared: "If we want to know how people feel, what they experience and what they remember, what their emotions and motives are like, and the reasons for acting as they do, why not ask them?" (p. 37). Wiggins observed that, "...regardless of the theoretical considerations which guide scale construction or the mathematical elegance of item-analytic procedures, the practical utility of a test must be assessed in terms of the number and magnitude of its correlations with nontest criterion measures" (1973, p. 413).

Therefore, if, when compared on criterion measures, a single item asking for a self-rating on a dimension of personality were as valid as a lengthy multiple item scale, why not use the self-rating instead? A single item self-rating is certainly more economical than a multiple item scale, and may "...reduce the

*Larry C. Bernard, Ph.D.; Psychology Department; Loyola Marymount University; One LMU Drive; Los Angeles, CA 90045; lbernard@lmu.edu (email).

fatigue, frustration, and boredom..." that can occur from answering similar, repetitive questions (Robins, Hendin, & Trzesniewski, 2001, p. 152).

However, longer instruments may have better psychometric properties than shorter ones, while single items present potential measurement issues. First, is the issue of reliability. Some of the customary methods of assessing reliability, such as Cronbach's α (internal consistency) or factor analysis (homogeneity and unidimensionality), are not applicable to single items, but others, such as test-retest (stability) are. Second, is the issue of validity. Just how valid could domain self-ratings be, when compared to a multiple item scale?

Serious questions have been raised previously about the presumed superior validity of the standard, lengthy, objective questionnaire method (e.g., Ashton & Goldberg, 1973; Burisch, 1978; Hamilton, 1971; Hase & Goldberg, 1967; Harris, 1980; Kaufman & Murphy, 1981; Lamont, 1983). The questioning seems to have culminated in Burisch's (1984a) review of personality inventory construction methods:

"...(very short) self-rating scales narrowly but consistently outdo questionnaire scales in terms of validity and are clearly superior in terms of communicability and economy. There may not be many situations in which the widespread preference for questionnaires is justified. It is concluded that the more commonsensical approaches to personality measurement have a lot to offer" (Burisch, 1984, p. 214).

Prior research has supported this commonsensical or self-rating approach to personality assessment. Early studies by Goldberg and his associates (e.g., Ashton & Goldberg, 1973; Goldberg, 1972; Hase & Goldberg, 1967) found that different personality scale construction strategies produced relatively little difference in validity. Burisch (1978) replicated and extended these studies, comparing various construction strategies for personality questionnaires and taking into account weaknesses in the designs of Goldberg's studies. Burisch (1984b) also evaluated various strategies of scale construction and compared the validity of scales on characteristics such as: length, simplicity versus sophistication, and self-rating. His results (though not entirely consistent) generally supported the following conclusions: Differences in scale length and sophistication did not affect validity. Furthermore, very short self-ratings (such as a rating scale consisting of a trait label and pairs of opposite adjectives) were as valid as standard, lengthy questionnaires.

More recently, Burisch (1997) empirically investigated the issue of optimum scale length. Burisch was reacting to Spearman's (1910) "...assertion that lengthening a test will improve its reliability, and indirectly, its validity..." (1997, p. 304). Others (Bell & Lumsden, 1980; Ulrich, 1985) had shown previously that shortened tests do not necessarily lose validity. Bell & Lumsden (1980) used two 100-item scholastic aptitude tests in economics and English comprehension to demonstrate that both could be shortened by 40 items each without loss of validity. They concluded with the rather stunning statement: "...all tests could be reduced by more than 60% without appreciable decreases in validity" (p. 165).

Burisch (1997) was interested in testing this assertion with the measurement of personality constructs. He obtained three samples for cross-validation purposes and used an obsolete German-language inventory that was comprised of multiple item scales measuring these personality constructs: somatic lability, aggressiveness, depressiveness, excitability, gregariousness, calmness, dominance, and inhibition. Each construct was measured by a scale that contained between 20 and 34 items. Using MAXVAL, a specialized statistical program (Ulrich & Giray, 1984; see Burisch, 1997 for a description), Burisch was able to demonstrate that there was "...hardly any loss of cross validity" on double cross-validation using "...extremely short scales of two to four items each.." (p. 303).

More recently, Gosling, Rentfrow, and Swann (2003) constructed and tested two very brief 5 and 10-item measures of the Big-Five personality domains. The Big-Five model posits five broad individual differences factors of personality, each comprised of several lower-order facets. Costa and McCrae's (1992) *NEO Personality Inventory-Revised* (NEO PI-R; Costa & McCrae, 1992) is one of the most popular contemporary instruments designed to measure the Big Five domains (Cohen & Swerdlik, 2002). The NEO PI-R consists of 240-items and takes approximately 45 minutes to complete. Gosling et al. were able to demonstrate reasonable levels of reliability and validity, including convergence between self and observer ratings, for their 10-item per domain measure, which takes as little as 10 minutes to complete.

Returning to Allport's position, and considering the results obtained by Burisch, Goldberg, and particularly Gosling et al., we wondered if the commonsensical approach could be extended a little further. Could we not simply ask people *once* to tell us – that is, provide a single, self-rating of – their standing on a description of each of the Big-Five domains? Gosling et al. used 10-items per domain while Burisch's results had indicated that two to four item scales might be valid. Would it be too much to ask, "is asking *once* enough?"

In an attempt to answer this question, observer ratings of participants on the Big-Five domains as represented in the NEO PI-R – Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness – were used as the criterion. The amount of variance accounted for in the criterion by participants' scores on the NEO PI-R scales and by participants' self-ratings on a single-item description of each domain were compared. We selected the NEO PI-R because its authors used a rational-empirical strategy of construction, with observer ratings serving as "one of the cornerstones of the validation program" (Costa & McCrae, 1992, p. 48). In validating the NEO PI-R, participants' scores were correlated with the criterion of observer ratings of the participants made by their spouses or peers nominated by the participants (e.g., McCrae, 1982; McCrae & Costa, 1987). Likewise, in the present study, observer NEO PI-R ratings of participants on the Big Five domains would also provide the external criterion.

Method

Participants

One hundred forty-nine participants were recruited through and received partial course credit in a university subject pool. Ninety-seven (65%) of the participants were women and 52 (35%) were men. Their mean age was 19 ($SD = 1.1$; range 18 to 25 years). The ethnic self-identification of the sample was consistent with that of the institution's general student body: 74 (50%) were Euro-American/Caucasian, 23 (15%) were "mixed" (more than one category), 20 (13%) were Latino/a-American, 16 (11%) were Asian-Pacific American, 10 (7%) were African American, and 6 (4%) did not report ethnicity. Participants were treated in accord with the ethical standards of the American Psychological Association and all data were collected anonymously.

Raters. Participants recruited acquaintances to rate them on the NEO PI-R Form R. Raters were asked to report their relationship with the participant they were rating. Of the 149 raters, 147 (2 demographic forms were missing) reported. Their mean age was 22 ($SD = 8.8$; range 16 to 66 years). One hundred nineteen (80%) of the raters reported the person they were rating was a "friend," 25 (17%) a "relative," and 3 (2%) indicated "other." The mean length of time raters reported knowing the person being rated was 57 months, and the median was 24 months. Raters were also asked how well they knew the participant they were rating. Ninety-four (63%) reported "very well," 44 (30%) reported "moderately well," and 9 (6%) reported "somewhat well." None reported "not well." The ethnic self-identification of raters was consistent with that of the participants: 70 (47%) were Euro-American/Caucasian, 24 (16%) were Latino/a-American, 20 (13%) were Asian-Pacific American, 17 (11%) were "mixed," 10 (7%) were African American, and 6 (4%) did not report. Raters were treated in accord with the ethical standards of the American Psychological Association and all data were collected anonymously.

Materials

The *NEO Personality Inventory - Revised* (NEO PI-R, Costa & McCrae, 1992) is a 240-item measure of the "Big Five" personality domains: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Participants completed Form S and raters completed Form R. Both forms have identical items, with the exception that "he" or "she" in Form R are substituted for "I" in Form S. The *NEO PI-R Professional Manual* describes Neuroticism as contrasting adjustment with maladjustment; Extraversion as sociability, liking people, preferring large groups, assertiveness, activity, and talkativeness; Openness as openness to experience, general curiosity; Agreeableness as altruistic, sympathetic toward and eager to help others; and Conscientiousness as self-control, planning, organizing, purposeful, and scrupulous (Costa & McCrae, 1992).

The scales were constructed from a rationally guided factor analytic process. Each domain scale is comprised of six facets, subscales that tap more

circumscribed traits related to the domain. The NEO PI-R Manual (Costa & McCrae, 1992) reports good internal consistency for the domain scales (Cronbach's $\alpha = .86$ to $.92$ for Form S and $.89$ to $.95$ for Form R) and good re-test reliability, ranging from $r = .87$ to $.91$ for earlier Form S versions of the Neuroticism, Extraversion, and Openness scales, and from $r = .75$ to $.83$ for shorter versions of all five domain scales. Form S stability coefficients from a seven-year longitudinal study of peer ratings ranged from $.63$ to $.81$ for the five domain scales. The Manual also reports several studies attesting to the NEO PI-R's convergent, divergent, and criterion validity. In particular, the authors have made extensive use of observer ratings as objective validity criteria, which is the primary reason they were used as the objective criterion in the present study.

Your Personality Form. This is a specially-constructed form on which participants were asked to provide self-ratings on five single items that described the NEO PI-R domains. The NEO PI-R Manual includes a section called "Your NEO Summary," that offers language appropriate for feedback directly to lay respondents, including a verbal description for the high, intermediate, and low point of each of the five domains. The domain scale names were not used on the Your Personality Form. Instead, the domains were identified simply as "Trait #1, Trait #2....Trait #5." For each trait, the three verbal descriptions of its high, intermediate, and low point were provided along a Likert scale that ranged from 1 to 7. Participants were asked to circle the number that indicated where they fell on the descriptive continuum of each trait. For purposes of the present study, these scores are identified as the "domain self-ratings" (DSRs). Costa and McCrae (1992) report a study in which 3-point ratings on these descriptions yielded correlations between $r = .24$ to $.60$, concluding these values support the validity of using them for feedback but not as a substitute for the NEO PI-R scales. Use of a 7-point might increase the correlations for the purposes of the present study.

Procedure

Participants self-selected and signed-up for groups of approximately 10 to 15 at various sessions during a 16-week semester. The procedure was conducted in two parts. Research assistants met the group sessions, provided oral and written instructions for both Parts 1 and 2, and obtained participants' informed consent. In Part 1, participants completed the NEO PI-R Form S and the Your Personality Form (in randomized counterbalanced order) and a brief demographic questionnaire. Participants then received credit for Part 1.

In Part 2, participants were instructed to recruit a rater who knew the participant for at least 6 months to rate them on the NEO PI-R Form R. The rater could be a friend or relative. Participants gave raters a large envelope covered by an informed consent form, which the raters were to sign before opening the envelope. Inside, raters found additional instructions to complete the enclosed items (the NEO PI-R Form R and brief demographic questionnaire), seal the envelope in a manner that would make it tamper-evident, and give it back to the

participant. The participants had been instructed during Part 1 to return the sealed envelope in an un-tampered condition along with the rater's informed consent form for full credit.

Of 165 subjects participating in Part 1, 152 (92%) completed Part 2. Materials from Part 1 and Part 2 were then matched by a pre-coded sequence number which appeared on all materials (except for the informed consent forms which were kept separately). In an attempt to verify that the participants did not also complete the raters' Form R, the returned envelopes were inspected for evidence of tampering and participants' and raters' materials were inspected for similarity of handwriting style. Two envelopes were returned unsealed and apparently tampered with, and one envelope was returned with multiple answers marked for some items. Materials from these three cases were dropped from the analysis, yielding the final total of 149 completed participant and rater materials.

Results

In order to help determine the generalizability of these results, we first analyzed the present sample's similarity with the NEO PI-R College-Age sample as reported in the Manual (Costa & McCrae, 1992). Mean NEO PI-R scale scores of the present sample were compared to the NEO PI-R Form S College-Age Sample. Results are depicted in Table 1. The NEO PI-R scales are theoretically independent. They were developed factor analytically, with rotations to orthogonal Varimax criteria. Resulting intercorrelations among Form S domain scales range from $r = .02$ to $.53$, median = $.21$ (Costa & McCrae, 1992). Therefore, each domain was analyzed independently. The NEO PI-R Manual also provides separate norms for men and women. Therefore, the significance of differences in means between the present sample and the College-Age Sample was tested by computing one-sample t -tests for each of the five domains separately for men and women. A Bonferroni correction of $\alpha = .01$ [$\alpha = .05 \div 5 = .01$] was used as the critical level in order to hold Type I error constant across the five domains for each sex (similar corrections were applied throughout). With that criterion, only one mean difference was significant. Women in the present sample had a higher Neuroticism mean score (106.4, $SD = 22.1$) than women in the NEO PI-R College Age Sample (99.8, $SD = 20.9$) [$t(96) = 2.96, p = .004$].

Table 2 depicts the means, standard deviations, and minimum and maximum values on the domain self-ratings (DSRs) and Form S and Form R scores. These data indicate there is no restriction in range in this sample.

Zero-order correlations were computed between the DSRs, Form S scores, and Form R scores and are presented in Table 3. These correlations help in assessing the convergent and divergent validity of the measures. The first five columns of Table 3 present the correlations between each DSR and the self-ratings on the single items for the other four domains. None of the correlations were significant ($p > .005$; alpha levels are Bonferroni-corrected), supporting the single items' divergent validity.

Table 1
Means and (Standard Deviations) of the Present Sample and the College Normative Sample on the NEO PI-R's Five Domain Scales by Sex

Domain	Present Sample	College Normative Sample
	Men	
Neuroticism	92.3 (22.1)	90.5 (22.1)
Extraversion	117.2 (18.9)	116.7 (18.3)
Openness	120.4 (19.7)	113.9 (18.5)
Agreeableness	106.7 (14.3)	107.4 (16.2)
Conscientiousness	106.7 (20.8)	113.5 (22.0)
	Women	
Neuroticism*	106.4 (22.1)	99.8 (20.9)
Extraversion	126.0 (20.5)	123.9 (17.7)
Openness	123.4 (18.4)	118.6 (17.1)
Agreeableness	113.0 (25.1)	117.2 (15.7)
Conscientiousness	113.2 (23.3)	115.1 (20.6)

Note. Present sample $N = 98$ women and 52 men; NEO PI-R College-Age sample $N = 148$ men and 241 women. Bonferroni corrected $\alpha = .01$.

* $p < .004$

The middle five columns of Table 3 contain the correlations between each Form S domain scale and the other four Form S domain scales. As with the DSRs, none of the correlations were significant ($p > .005$), also supporting Form S scores' divergent validity. The middle five columns of Table 3 also contain the correlations between the DSRs and the Form S scale scores. The single boxed diagonal values contain the (convergent) correlations between DSRs and Form S scores for the same domain. The off-diagonal values are the (divergent) correlations between the DSRs and Form S scores on other domains. None of the off-diagonal correlations were significant ($p > .003$) and all of the on-diagonal correlations were significant ($p < .003$).

Table 2
Means, Standard Deviations, Minimum and Maximum Values on Single Items, and NEO PI-R Form S and Form R Scores

	Mean	Standard Deviation	Minimum	Maximum
<i>Single Item</i>				
Neuroticism	3.94	1.35	1	7
Extraversion	3.57	1.34	1	7
Openness	3.34	1.31	1	7
Agreeableness	3.49	1.24	1	7
Conscientiousness	3.57	1.52	1	7
<i>NEO PI-R Form S</i>				
Neuroticism	101.46	23.02	53	175
Extraversion	122.95	20.35	74	172
Openness	122.40	18.89	66	174
Agreeableness	110.74	19.77	39	156
Conscientiousness	110.95	22.63	52	163
<i>NEO PI-R Form R</i>				
Neuroticism	92.73	22.14	36	162
Extraversion	118.91	20.58	59	164
Openness	110.67	17.02	71	155
Agreeableness	110.74	24.96	15	169
Conscientiousness	111.13	25.31	52	173

The last five columns of Table 3 present the correlations between Form S scores and the single item self ratings and Form R scores, and directly address the research question. The double boxed diagonal values contain the (convergent) correlations between DSRs and Form S scores and Form R scores for the same domain. The off-diagonal values are the (divergent) correlations between the DSRs and Form S scores and Form R scores on other domains. All on-diagonal correlations were significant ($p < .003$) while none of the off-diagonal correlations were significant ($p > .003$).

Hierarchical multiple regression analysis was used to assess the relative amount of variance in Form R scores accounted for by the DSRs and Form S scores. Five pairs of multiple regressions were calculated, one for each Form R domain score. In the first of each pair of regressions, the DSR was entered first,

Table 3
Zero-Order Correlations of NEO-PI-R Form S Scores and Single Items with each other and with NEO PI-R Form R Scores

	NEO PI-R Form S										NEO PI-R Form R				
	E	O	A	C	N	E	O	A	C	N	E	O	A	C	
N															
NEO PI-R Form S															
Single Item	--	-21	-11	05	08	60*	-16	-07	04	-04	41*	-08	-07	-06	05
E															
NEO PI-R Form S															
Single Item	--	20	-15	-04	-04	-21	67*	00	-04	04	-05	56*	16	-02	-01
O															
NEO PI-R Form S															
Single Item	--	08	-10	-10	-15	-15	23	46*	06	22	05	12	29*	15	13
A															
NEO PI-R Form S															
Single Item	--	--	-05	-04	-04	-04	-19	08	52*	-04	03	08	02	31*	-08
C															
NEO PI-R Form S															
Single Item	--	04	10	-14	-14	07	73*	04	14	-15	-15	-15	-15	-15	57*

Table 3 Note. $N = 149$. Decimals omitted. N = Neuroticism, E = Extraversion, O = Openness, A = Agreeableness, and C = Conscientiousness.

A Bonferroni correction of $\alpha = .003$ [$\alpha = .05 \div 15 = .0033$] was used as the critical level in order to hold Type I error constant across all 15 comparisons.

* $p < .0005$

Table 4

Summary of Hierarchical Regression Analyses for Variables Predicting the Five NEO PI-R Domains and Magnitude of R^2 Change ($N = 149$)

NEO PI Domain	Order of Entry	R	R^2	R^2 Change	B	$SE B$	β
N	1. Single Item	.41	.17		4.151	1.51	.25*
	2. NEO PI-R Form S	.46	.21	.04*	.25	.09	.26*
	1. NEO PI-R Form S	.41	.17		.25	.09	.26*
	2. Single Item	.46	.21	.04*	4.15	1.51	.25*
E	1. Single Item	.46	.21		2.25	1.39	.15
	2. NEO PI-R Form S	.58	.33	.12*	.47	.09	.46*
	1. NEO PI-R Form S	.56	.32		.47	.09	.46*
	2. Single Item	.58	.33	.01	2.25	1.39	.15
O	1. Single Item	.29	.08		.20	.98	.02
	2. NEO PI-R Form S	.60	.36	.28*	.53	.07	.59*
	1. NEO PI-R Form S	.60	.35		.53	.07	.59*
	2. Single Item	.60	.36	.01	.20	.98	.02
A	1. Single Item	.31	.09		1.73	1.72	.09
	2. NEO PI-R Form S	.48	.23	.14*	.54	.11	.42*
	1. NEO PI-R Form S	.47	.22		.54	.11	.42*
	2. Single Item	.48	.23	.01	1.73	1.72	.09
C	1. Single Item	.57	.32		5.75	1.59	.34*
	2. NEO PI-R Form S	.61	.37	.05*	.35	.11	.31*
	1. NEO PI-R Form S	.56	.32		.35	.11	.31*
	2. Single Item	.61	.37	.05*	5.75	1.59	.34*

* $p < .01$

Table 5
Zero-Order Correlations of NEO PI-R Form S Facets with Domain Self-Ratings (DSRs), and NEO PI-R Form S and Form R Domain Scores

Form S Facet	Neuroticism			Extraversion			Openness			Agreeableness			Conscientiousness		
	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R
Neuroticism															
-Anxiety	53	76	39	-13	-06	01	-12	-05	-02	12	-01	01	13	04	12
-Angry Hostility	36	66	24	-03	-12	-03	-15	-17	-09	-38	-53	-25	10	-12	06
-Depression	50	81	34	-29	-28	-11	10	-04	-02	11	-08	02	02	-22	05
-Self-consciousness	36	72	18	-26	-27	-08	-16	-16	-12	12	05	04	04	-10	14
-Impulsiveness	27	62	33	02	01	-02	00	09	03	-21	-31	-16	-14	-38	-20
-Vulnerability	60	80	33	-20	-21	-13	-10	-11	-15	05	-09	04	02	-24	04
Extraversion															
-Warmth	-02	-14	08	40	69	37	17	18	-06	17	44	23	00	13	-09
-Gregariousness	-13	-09	-07	66	79	41	16	08	-06	-05	10	03	04	07	-03
-Assertiveness	-14	-16	-09	53	68	47	08	08	18	40	-29	-24	19	19	16
-Activity	-06	-18	08	41	67	48	10	14	14	-20	-03	00	25	35	13
-Excitement Seeking	-12	-04	00	45	65	24	19	28	09	-14	-06	-06	-08	-07	-15
-Positive Emotions	-18	-28	-03	27	65	35	26	26	02	-10	18	04	01	15	-08
Openness															
-Fantasy	-03	06	10	00	20	03	28	68	31	09	-11	-10	23	-28	-26
-Aesthetics	08	01	11	02	09	08	37	76	28	06	08	-15	01	01	08
-Feelings	17	13	16	08	30	19	14	56	20	08	-04	-16	12	-01	01
-Actions	-18	-24	03	15	22	15	52	62	42	02	-01	-08	09	-03	-03
-Ideas	-21	-25	-06	04	09	09	21	66	38	07	08	-05	02	08	00
-Values	-10	-13	-01	-13	04	-02	24	46	22	10	10	03	17	-08	-11
Agreeableness															
-Trust	-09	-35	00	20	22	26	13	05	00	28	65	25	00	14	-05
-Straightforwardness	07	-16	09	-10	-01	02	-08	04	-02	32	71	31	18	33	19
-Altruism	-04	-22	10	10	31	25	11	11	-06	32	68	33	11	30	11
-Compliance	05	-20	-10	-06	-06	01	02	-05	-06	34	74	30	06	04	-01
-Modesty	19	23	21	-13	-11	-03	-07	-06	-08	22	45	26	03	-02	11
-Tender-Mindedness	01	-04	07	-04	07	-05	20	16	-12	29	50	28	-16	-14	-07

Form S Facet	Neuroticism			Extraversion			Openness			Agreeableness			Conscientiousness		
	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R
Conscientiousness															
-Competence	-16	-33	-15	08	24	12	-18	09	-01	-05	09	01	46	74	33
-Order	03	-07	00	-02	10	10	-16	-02	00	-10	06	02	66	73	50
-Dutifulness	06	-09	06	03	18	14	-14	-06	-12	-01	31	09	46	74	39
-Achievement Striving	01	-05	06	16	37	21	10	-04	-01	-16	-01	-09	70	76	43
-Self-Discipline	-10	-38	-10	12	21	18	16	-06	00	-04	16	02	58	84	47
-Deliberation	-01	-10	00	-18	-14	-08	-26	-24	-20	16	14	14	38	67	37

Note. N = 149. NEO PI-R Form S and R (Costa & McCrae, 1992). Convergent correlations shown in bold typeface.

Table 6
Zero-Order Correlations of NEO PI-R Form R Facets with Domain Self-Ratings (DSRs), and NEO PI-R Form S and Form R Domain Scores

Form R Facet	Neuroticism			Extraversion			Openness			Agreeableness			Conscientiousness		
	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R
Neuroticism															
-Anxiety	48	37	73	-02	08	-03	01	06	-12	10	16	02	09	07	03
-Angry Hostility	15	29	59	08	04	-09	08	14	-03	-14	-26	-64	15	06	-03
-Depression	27	36	79	-14	-08	-22	07	10	-09	10	10	-04	02	-06	-16
-Self-consciousness	28	21	66	-09	-08	-17	-03	-09	-26	05	20	12	07	11	-1
-Impulsiveness	12	18	60	07	09	03	10	25	11	-13	-13	-24	-12	-20	-39
-Vulnerability	40	30	80	-09	-10	-23	00	-09	-32	14	18	00	08	-13	-35
Extraversion															
-Warmth	03	02	-16	18	37	68	02	09	24	06	24	55	04	01	09
-Gregariousness	-02	-10	-04	53	62	79	10	12	08	02	20	09	10	11	-03
-Assertiveness	-13	-06	-26	37	29	65	11	-02	27	-27	-16	-34	17	12	30
-Activity	05	02	04	27	32	66	05	00	25	-13	04	-05	16	15	28
-Excitement Seeking	-15	-15	-16	29	38	70	13	20	40	-06	-03	-03	07	02	-01
-Positive Emotions	05	-05	-15	24	37	72	07	17	34	05	18	40	11	17	23

Form R Facet	Neuroticism			Extraversion			Openness			Agreeableness			Conscientiousness		
	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R	DSR	S	R
Openness															
-Fantasy	-14	-04	07	14	19	28	27	39	55	03	-18	-04	-30	-29	-30
-Aesthetics	06	00	-13	-17	-10	10	00	14	66	06	-03	08	10	09	24
-Feelings	17	14	13	07	15	34	09	39	50	-16	-19	-08	04	01	16
-Actions	-10	-11	-11	18	12	30	21	14	42	07	15	11	17	-12	-03
-Ideas	-21	-16	-38	05	04	21	10	35	65	03	03	01	03	00	25
-Values	-13	-15	-08	01	07	17	21	38	48	02	-04	09	19	-11	-09
Agreeableness															
-Trust	-06	-07	-24	01	14	32	-04	-18	07	16	41	78	05	02	07
-Straightforwardness	14	-08	-17	-13	-07	-04	-15	-12	-01	27	37	80	00	13	34
-Altruism	08	-03	-27	-06	07	38	-10	-07	17	18	36	82	04	09	28
-Compliance	10	-07	-16	-19	-12	-06	-19	-18	-10	35	36	81	07	02	03
-Modesty	02	05	-02	-15	-07	-07	-15	-13	-03	21	31	76	-19	-08	-07
-Tender-Mindedness	01	-04	-06	15	-01	15	04	11	22	26	31	66	01	03	16
Conscientiousness															
-Competence	-05	-03	-41	01	00	25	-11	05	27	-08	00	11	35	38	79
-Order	08	07	02	-05	01	04	00	-04	08	-06	07	-01	54	49	76
-Dutifulness	09	08	-21	-05	-01	22	-15	-06	18	-05	12	31	43	44	86
-Achievement Striving	11	08	-09	04	09	33	-07	-01	20	-12	07	-03	53	48	80
-Self-Discipline	10	06	-19	-07	-01	20	16	-07	11	-06	05	10	53	51	90
-Deliberation	00	-01	-30	-04	-09	02	-17	-20	-11	-05	01	23	38	42	76

followed by the Form S score. The order was reversed in the second of each pair of regressions. The results of the change in R^2 for all 10 multiple regression analyses are depicted in Table 4. Regardless of the order of entry, the amount of R^2 change was identical and significant ($p < .01$) for the Neuroticism and Conscientiousness domains, suggesting the DSRs and Form S scores performed equivalently. For Extraversion, Openness, and Agreeableness, Form S scores added significantly to the amount of R^2 change ($p < .01$), whereas the DSR did not ($p > .05$). Form S scores for the Extraversion, Openness, and Agreeableness domains accounted for more variance in Form R scores than the DSRs.

One advantage of the lengthier NEO PI-R is its ability to measure six facets for each Big-Five domain. Table 5 depicts the zero-order correlations of each of

the NEO PI-R Form S Facets with the DSRs, Form S, and Form R domain scores, while Table 6 depicts the zero-order correlations of each of the NEO PI-R Form R Facets with the DSRs, Form S, and Form R domain scores. These tables allow an appraisal of the convergent and divergent validities of the various facets with the domains as a function of the method of measurement. As would be expected, Form S facet scores correlate higher with Form S domain scores than they do with Form R domain scores or DSRs, and Form R facet scores correlate higher with Form R domain scores than they do with Form S domain scores or DSRs. The most germane comparison is between the Form R facet scores and the Form S domain scores and DSRs in Table 6. These results are generally consistent with those of the multiple regression analyses. More Form S domain scores correlate more highly with Form R Extraversion, Openness, and Agreeableness domain scores than DSRs do, while the correlations of Form S domain scores and DSRs are more similar with Form R Neuroticism and Conscientiousness.

Discussion

The present study used domain self-ratings to push investigation of the issue of scale size and validity to its limit for the Big-Five domains. Burisch (1997) had previously demonstrated two to four item scales can have adequate validity with various personality constructs, while Gosling et al. (2003) have shown that 10-item scales can have adequate validity measuring the Big-Five domains. We attempted to extend Burisch's (1997) and Gosling et al.'s findings by examining the validity of single item self-ratings on descriptions of the Big-Five domains.

However, rather than a general finding that single item self-ratings of Big-Five personality domains is as valid as lengthier, multiple item scales, instead it appears that the validity of single item self-ratings of the Big-Five factors may be domain specific. Domain self-ratings predicted the criterion of observer ratings on Neuroticism and Conscientiousness as well as Form S scale scores. However, Form S scale scores added significantly to the prediction of Extraversion, Openness, and Agreeableness ratings, over and above domain self-ratings. The multiple item NEO PI-R does appear to offer advantages for the measurement of at least some Big-Five domains.

There are some limitations of the present study. Primarily, it used a university sample that may not reflect the broad range of the general population. As to the external validity of the present study with regard to college students, the means of this sample on the NEO domains, except in one case, were not significantly different from the NEO PI-R College Age sample, suggesting the present sample is much like the College Age Sample used to provide normative data for the NEO PI-R. Furthermore, the correlations between participants' scores and self-ratings were similar to those reported in the NEO PI-R Manual and used to establish concurrent validity for the NEO PI-R scales. A second limitation of the present study is that it only looked at one personality questionnaire and there are dozens in current use. However, the NEO PI-R is one of the most popular

instruments in use today. The present study's final limitation is methodological. Form S and Form R of the NEO PI-R contain identical response choices and identical items except for the pronouns used. This identity of method could tend to produce higher correlations between Form S and Form R than between simple self-ratings and Form R scores, but if this were true (and we do not know), it would have provided an advantage for Form S scores and, therefore, a more stringent test of our hypothesis.

Until now, a cumulative line of research (Ashton & Goldberg, 1973; Bell & Lumsden, 1980; Burisch, 1978, 1984a, 1984b, 1997; Hamilton, 1971; Hase & Goldberg, 1967; Harris, 1980; Kaufman & Murphy, 1981; Lamont, 1983; Ulrich, 1985) has suggested that valid measurement of personality can be accomplished with scales comprised of very few items, perhaps as few as two to four (Burisch, 1997). The present study suggests that even a single self-rating item may be sufficient for measuring at least two of the Big-Five personality domains: Neuroticism and Conscientiousness. However, the results of the present study also stand in contrast to this line of research, suggesting that lengthier scales may add significant validity for the measurement of three other Big-Five domains – Extraversion, Openness, Agreeableness – at least when compared to domain self-rating scales.

Generally, the pattern of correlations of the NEO PI-R Form R facet scores with the Form S scores and domain self-ratings coincided with the results overall. Yet, they point out the inability of the single-item domain self-ratings to measure the narrower aspects of personality represented in each domain's facets. If only a measure of the broader personality domains is required, then the single item self-ratings could possibly suffice for some domains. However, the facets may be better predictors of specific criteria than the broader Big-Five domains (Paunonen & Ashton, 2001), and if a more detailed description of personality is necessary, then a more complex instrument such as the NEO PI-R would be necessary. This issue of facet measurement suggests one intriguing direction for future research. Perhaps single-item self-rating descriptions might also prove to be valid indicators of some of the facets? The present study suggests it would be worth investigating for some of the facets of the Neuroticism and Conscientiousness domains.

The present study also points to the complexities inherent in the measurement of personality dimensions. Harkness, Tellegen, and Waller (1995) had shown higher correlations between self-report and informant ratings for Extroversion ("positive emotionality") than for Neuroticism ("negative emotionality") on the *Multidimensional Personality Questionnaire*. They also showed that ratings on Neuroticism were not uniform across different classes of raters (such as peers or parents). One implication is that different classes of raters have different access to information about a rated person's negative emotional behavior. While the present study does not directly address that issue, it also alerts us to different measurement issues dependent on the domain being measured: Single items may have adequate validity for some personality domains, while more lengthier scales

can add significantly to the validity of measurement for others. Future research should explore the scale-length issue on other personality domains further. In addition, future research might address the issue of an interaction between the personality construct being measured and the methodology used.

Finally, clinical and counseling psychologists often must balance their interest in efficient assessment procedures that are economical and facilitate client rapport against concerns about the supposed decreased validity of brief measures. These results suggest that domain self-ratings may provide an efficient and valid assessment of a client's standing on at least two Big Five personality factors – Neuroticism and Conscientiousness – which are important in a clinical setting. Domain self-ratings could also be repeated often, without unnecessarily straining rapport, to measure progress.

Author Note

We are grateful for the assistance of Amanda Garrett, Sara Kiser, Daniel Lehman, Adam Vaughn, and especially Charles Marriott in the conduct of this study.

References

- Allport, G. W. (1941). *The use of personal documents in psychological science*. NY: Social Science Research Council.
- Allport, G. W. (1955). *Becoming: Basic considerations for a psychology of personality*. New Haven, CN: Yale University Press.
- Ashton, S. G., & Goldberg, L. R. (1973). In response to Jackson's challenge: The comparative validity of personality scales constructed by the external (empirical) strategy and scales developed intuitively by experts, novices, and laymen. *Journal of Research in Personality, 7*, 1-20.
- Bell, R., & Lumsden, J. (1980). Test length and validity. *Applied Psychological Measurement, 4*, 165-170.
- Burisch, M. (1978). Construction strategies for multiscale personality inventories. *Applied psychological measurement, 2*, 97-111.
- Burisch, M. (1984a). Approaches to personality inventory construction: A comparison of merits. *American Psychologist, 39*, 214-227.
- Burisch, M. (1984b). You don't always get what you pay for: Measuring depression with short and simple versus long and sophisticated scales. *Journal of Research in Personality, 18*, 81-89.
- Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality, 11*, 303-315.
- Cohen, R. J., & Swerdlik, M. E. (2002). *Psychological testing and assessment: An introduction to test and measurement* (5th ed.). Boston: McGraw Hill.
- Costa, P. T., Jr., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources, Inc.
- Goldberg, L. R. (1972). Parameters of personality inventory construction and

- utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs*, 72-2.
- Gosling, S. D., Rentfrow, P. J., & Swann, Jr., W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504-528.
- Hamilton, D. L. (1971). A comparative study of five methods of assessing self-esteem, dominance, and dogmatism. *Educational and Psychological Measurement*, 31, 441-452.
- Harkness, A. R., Tellegen, A., & Waller, N. (1995). Differential convergence of self-report and informant data for Multidimensional Personality Questionnaire traits: Implications for the construct of negative emotionality. *Journal of Personality Assessment*, 64, 185-204.
- Harris, J. G., Jr. (1980). Nomovalidation and idiovalidation: A quest for the true personality profile. *American Psychologist*, 35, 729-744.
- Hase, H. D., & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin*, 67, 231-248.
- Kaufman, L., & Murphy, N. C. (1981). Validation through self and other ratings on dimensions of six nonstressful multi-scaled personality instruments. *Journal of Personality Assessment*, 45, 86-89.
- Lamont, D. J. (1983). A three dimensional test for white's effectance motive. *Journal of Personality Assessment*, 47, 91-99.
- McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, 43, 293-303.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, 81, 524-539.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem scale. *Personality and Social Psychology Bulletin*, 27, 151-161.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.
- Ulrich, R. (1985). Die Beziehung zwischen Testlänge und Validität für nicht-parallele Aufgaben: Verschiedene Methoden der Validitätsmaximierung [The relationship of test length and validity for non-parallel items: various methods of validity maximization]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 6, 32-45.
- Ulrich, R., & Giray, M. (1984). Improved algorithms and a Fortran program (MAXVAL) for selecting the most valid items from a pretested item pool. *Berichte aus dem Psychologischen Institut der Universität Tübingen*, 13.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Received 05/16/2004; Revision Received 09/03/2004; Accepted 09/03/2004

Copyright of Counseling & Clinical Psychology Journal is the property of Individual Differences Research Group and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.