The construct of agreeableness: Facet vs. item level analysis

Rebecca A. Newgent, Sang Min Lee, Kristin K. Higgins, and Sean W. Mulvenon
University of Arkansas
Joanie V. Connors
Fayetteville, Arkansas

Abstract
The Revised NEO Personality Inventory (NEO PI-R) was developed to operationalize the Five-Factor Model of Personality. Using correlational analysis and confirmatory and exploratory factor analysis, the present study investigates the facet structure of the domain of Agreeableness of the NEO-PI-R at the facet and item level to assess which is a more appropriate level of analysis for interpretation. While the six Agreeableness facets fit the domain, the results of confirmatory factor analysis indicate that the original six facet model at the item level did not fit the data well. Exploratory factor analysis indicates that only 35 of the 48 items associated with the Agreeableness domain had salient loadings on their intended facet. Results indicate that performing item-level factor analysis can produce misleading or un-interpretable results. Implications for research and assessment are discussed.

The Construct of Agreeableness: Facet vs. Item Level Analysis

The Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992a) was originally developed in 1978 as the NEO Inventory. The NEO Inventory consisted of three domain scales and 18 facet scales which measured Neuroticism, Extraversion, and Openness to Experience (Costa & McCrae, 1992b). Scales, consisting of 18 items, were added in 1985 to measure the domains of Agreeableness and Conscientiousness, and the inventory was renamed the NEO Personality Inventory. These new domain scales were inadequate in capturing the complete personality picture as the number of items was limited. Then, in 1990, facet scales were completed for Agreeableness and

Conscientiousness, with the addition of several items for each scale, and minor modifications made to the original items measuring Neuroticism, Extraversion, and Openness to Experience. "It is this version that we refer to as the Revised NEO Personality Inventory or NEO PI-R" (Costa & McCrae, 1992b, p. 40).

Personality researchers generally agree that the Five-Factor Model of Personality (FFM) represents the most basic dimensions underlying the traits of personality. The NEO PI-R was developed by Costa and McCrae (1992b) to operationalize the FFM. "The Revised NEO Personality Inventory (NEO PI-R) is a concise measure of the five major dimensions, or domains, of personality and some of the more important traits or facets that define each domain. Together, the 5 domain scales and 30 facet scales of the NEO PI-R allow a comprehensive assessment of adult personality" (Costa & McCrae, 1992b, p. 1).

The scales of the NEO PI-R were developed through use of both rational and factor analytic methods. According to Costa and McCrae (1992b), the NEO PI-R scales "have been the subject of intensive research conducted for 15 years on both clinical and normal adult samples" (p. 1). The majority of the psychometric research conducted on the current version has focused on the domains of the NEO PI-R, not the facets (i.e., Costa & McCrae, 1997; Furnham, 1996; Loehlin, McCrae, Costa, & John, 1998; McCrae & John, 1992; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996).

Limited psychometric research has been conducted on the facets and items associated with the newer domains of Agreeableness and Conscientiousness since their original development in 1985 (i.e., Costa, McCrae, & Dye, 1991; McCrae & Costa, 1992; Piedmont & Weinstein, 1993). In 1991, Costa, et al. conducted separate factor analyses (varimax rotation) of the Agreeableness and Conscientiousness domains ($N = 1,539$). An Orthogonal Procrustes analysis resulted in 41 of the 48 items of both the Agreeableness and Conscientiousness domain scales loading on their intended factor. This study, however, did not include item analysis in relation to the facet level. In 1992, McCrae and Costa tested for valid specific variance of the 30 facet scales of the NEO PI-R. Results of principal component analyses for self-reports and peer ratings ($N = 250$) and spouse ratings ($N = 68$) on the 30 facet scales indicated a varimax-rotated five-factor solution consistent with the hypothesized structure of the facets. Although they reported how many facets load on their intended domain, similar to the previous study, this study did not include item analysis in relation to the facet level. Piedmont and Weinstein (1993) performed a psychometric evaluation of

the new scales of Agreeableness and Conscientiousness ($N = 236$) using principal components analysis with varimax rotation (5 factors were extracted and rotated). The facet scales resulted in substantial loadings on their intended factor with none of the Agreeableness or Conscientiousness facet scales having significant secondary loadings, supporting the construct validity of these facets. Again, this study reported how many facets load on their intended domain, but did not include item analysis in relation to the facet level.

While there is broad consensus regarding the five-factors, this consensus does not extend to the facet traits (Costa et al., 1991). This paper takes a closer look at the domain of Agreeableness and the facets and items that currently define it through a variety of statistical procedures. The purpose of this paper is to provide further psychometric evidence of validity through evaluation of the items and facets associated with the Agreeableness domain and establish if facet or item level analysis is more appropriate for interpretation.

Method

Participants

Two hundred forty ($N = 240$) participants ranging in age from 18 to 65 ($M = 31.54$, $SD = 11.20$) participated in this study. Participants consisted of 168 females (70%) and 72 males (30%). Two hundred seven (86%) were Caucasian and 33 (14%) were non-Caucasian. Seventy (29%) of the participants reported completing a high school education, 92 (38%) reported completing an associate or bachelor degree, and 78 (33%) reported completing a masters or doctoral degree.

Measure

NEO PI-R. The Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992a) consists of 240 items on a five-point scale ranging from strongly agree to strongly disagree. A substantial amount of research on the reliability and validity of the NEO PI-R has been conducted (Costa & McCrae, 1992b). Reviews state the NEO PI-R has established reliability from .86 to .95 for the domain level (Botwin, 1995; Juni, 1995). Internal consistency reliability coefficients for the facet scales range from .56 to .90 (Costa & McCrae, 1992b). Face validity appears high as the questions appear to be representative of personality characteristics. Additionally, consensual, construct, convergent, and

divergent validity have been established (Botwin, 1995; Costa & McCrae, 1992b; Juni, 1995).

Agreeableness domain. Forty-eight items and six facets comprise the domain of Agreeableness. "Agreeableness is primarily a dimension of interpersonal tendencies" (Costa & McCrae, 1992b, p. 15). The agreeable person is helpful, caring, cooperative, and has a basic belief that others will be helpful and caring in return. Individuals who score low on Agreeableness tend to be antagonistic, forceful, and disagreeable.

Agreeableness facet scales. The facet scales associated with Agreeableness include Trust, Straightforwardness, Altruism, Compliance, Modesty, and Tender-Mindedness (Costa & McCrae, 1992b). Individuals who score high in Trust tend to believe in the goodness of others, that they are well intentioned and honest (Costa & McCrae, 1992b). Individuals who score low in Trust tend to believe others are dishonest and have ulterior motives. Next, individuals who score high in Straightforwardness tend to be candid, trusting, and sincere. "Low scorers on this scale are more willing to manipulate others through flattery, craftiness, or deception" (Costa & McCrae, 1992b, p. 17). Individuals who score high on Altruism tend to be helpful, considerate, and intrinsically motivated while low scorers on this scale tend to be uninvolved and more self-interested. Fourth, individuals who score high on compliance tend to be forgiving, passive, and avoid conflict. "The low scorer is aggressive, prefers to compete rather than cooperate, and has no reluctance to express anger when necessary" (Costa & McCrae, 1992b, p. 18). Next, individuals who score high in Modesty are unassuming and humble while low scorers are overconfident and arrogant. The final facet, Tender-Mindedness, "measures attitudes of sympathy and concern for others. High scorers are moved by others' needs and emphasize the human side of social policies. Low scorers are more hardheaded and less moved by appeals to pity" (Costa & McCrae, 1992b, p. 18).

Procedure

An accessible sample of 240 participants was recruited on a voluntary basis from Colorado, New York, North Carolina, and Ohio from the general community and university programs. Each participant was administered the NEO PI-R and general demographic data was collected. Standardized written instructions for the NEO PI-R were provided. The researchers obtained written consent prior to test administration and these forms were collected and kept

separate to ensure anonymity of the information. Only the results of those participants who completed the NEO PI-R in its entirety were utilized.

Results

Facet Analysis

Overall, transformed *T*-Scores on the Agreeableness domain ranged from 20 to 80 with a mean of 47.57 and a standard deviation of 12.27. Table 1 presents descriptive statistics and alpha coefficients, separately by gender, for the Agreeableness domain and facet scales. The results of a series of *t*-tests indicate there are no statistically significant differences between male and female on the mean scores of all six facets ($p > .05$). Overall alphas for the Agreeableness facets ranged from .74 to .77. Alphas ranged from .73 to .77 for females on the Agreeableness facets and from .75 to .79 for males. Intercorrelations of the Agreeableness domain and its' facets are reported in Table 2. As can be seen, all the Agreeableness facets are significantly related to each other, with correlations ranging from .26 to .51 ($p < .01$). Correlations between the Agreeableness facets and the Agreeableness domain ranged from .57 to .62 ($p < .01$).

Table 1

Descriptive Statistics and Alpha Coefficients of the Agreeableness Domain and Facet Scales

Agreeableness Facet Scales	Overall Alpha[a]	Females[b]			Males[c]		
		M	*SD*	Alpha*	*M*	*SD*	Alpha*
Trust	0.75	47.59	12.53	0.75	47.65	9.62	0.75
Straightforwardness	0.74	46.29	11.39	0.73	45.87	10.88	0.76
Altruism	0.75	50.26	11.61	0.73	50.51	11.43	0.79
Compliance	0.74	46.00	10.84	0.74	46.14	10.61	0.77
Modesty	0.77	47.11	11.34	0.77	49.05	9.53	0.76
Tender-Mindedness	0.75	50.44	10.90	0.74	49.17	10.54	0.77
Agreeableness	0.78	46.83	12.18	0.78	49.29	12.41	0.80

Note:[a]*N* = 240. [b]*n* = 168. [c]*n* = 72. *Cronbach coefficient alpha with deleted variable. Cronbach coefficient alpha at the item level is .89.

Item Analysis

To test the fit of the proposed factor structure at the item level of the Agreeableness domain and associated facets, a confirmatory factor analysis was conducted with each participant's 48-item responses as input. The proposed model was designed as an oblique factor structure as the six facets of Agreeableness are significantly related to each other. An oblique model permits the factors to become somewhat correlated with each other in order to arrive at a clearer definition of simple structure. Results indicate an unacceptable fit of the proposed model of Agreeableness. The goodness of fit index (GFI) was only .74, and the adjusted goodness of fit index (AGFI) yielded only .72. Further, the root mean of square error of approximation (RMSEA) was .07, indicating an unsatisfactory fit to the sample data. Regarding the parameter standardized estimates, results show that twelve out of forty-eight item loading were below .40, among which six items were below .30. These results indicate that parameter estimates were not completely satisfactory for any of the six facet latent variables; the pattern of loadings did not match the a priori structure.

Table 2

Intercorrelations of the Agreeableness Domain and Facets

	A1	A2	A3	A4	A5	A6	Agreeableness
Trust (A1)	-	0.35	0.36	0.51	0.26	0.36	0.61
Straightforwardness (A2)		-	0.43	0.49	0.34	0.35	0.61
Altruism (A3)			-	0.33	0.34	0.42	0.60
Compliance (A4)				-	0.27	0.34	0.62
Modesty (A5)					-	0.42	0.57
Tender-Mindedness (A6)						-	0.62
Agreeableness							-

Note: $p < .01$ for all correlations.

Bagozzi and Heatherton (1994) report that an unsatisfactory fit is expected in a model which has more than four or five items per factor. The model of the present study has eight items per factor. Little, Cunningham, Shahar, and Widaman (2002) suggest that item parceling can be advantageous over using the original items because solutions from item-level data are less likely to yield stable solutions than solutions from parcels of items. Random methods of combining items can be used to create three item parcels per factor. Results

indicate a fit that was superior to that of the original proposed model due to item parceling procedure (GFI = .89; AGFI = .87; RMSEA = .06). The range of fit statistics, however, taken collectively, still indicates an unacceptable fit.

To further exam the underlying factor structure of the Agreeableness scale, an exploratory factor analysis was conducted. There were 14 factors with eigenvalues greater than one (Kaiser's criterion or Kaiser-Guttman test). Examination of the 'scree plot,' a plot with eigenvalues on the vertical axis and factor numbers (i.e., the first factor extracted, second, third, fourth and so on) on the horizontal axis, exhibited a very large first factor loading (yielded 17.3% of the total variance) with possible minor contributions from five additional factors (second through sixth factors contributed incremental amounts of 6.2%, 5.5%, 4.6%, 3.8%, and 3.5%). Along with the original theoretical framework (six-facet model), a six-factor solution was selected.

Results are presented in the order they appear in the assessment. An oblique promax rotation was specified for the 48 items that measure the Agreeableness domain. As can be seen in Table 3, when the 48 promax rotated items were examined, only 35 of the 48 items had salient loadings on their intended factor. While there were no multiple high loading items, four of the 13 items that did not significantly load on their intended facet loaded on another facet. All eight hypothesized items loaded on the facet of Trust in factor 1. One of the eight hypothesized items did not load on the facet of Altruism in factor 2. Next, two items did not load on Straightforwardness in factor 3 and three items did not load on Tendermindedness in factor 4. One of the eight hypothesized items did not load on the facet of Modesty in factor 5 and six items did not load on the facet of Compliance in factor 6.

Discussion

Facet Analysis

Regarding reliability information, Cronbach coefficient alphas for the facets associated with Agreeableness as well as the Agreeableness domain resulted in acceptable levels (.73 or higher) indicating support for the internal-consistency of these facets. In general, the alphas were slightly higher for males than females, with the exception of the Modesty facet (males = .76; females = .77). With the possible exception of the Altruism facet (males = .79; females = .73), no meaningful differences resulted between males and females in this sample. These results confirm the data presented by Costa and McCrae (1992b). This is contrary,

Table 3

Promax Rotated Factor Analysis of the Agreeableness Items

NEO PI-R Item		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
A1:	4	**.64**	.01	.18	-.06	-.07	-.08
	34	**.81**	-.17	-.05	-.05	.08	-.01
	64	**.71**	-.07	.18	-.03	-.07	-.06
	94	**.60**	.06	.05	.09	-.04	.07
	124	**.42**	.39	-.01	.00	-.07	-.19
	154	**.67**	.01	-.03	.09	-.07	.21
	184	**.60**	.12	-.16	.11	.07	.18
	14	**.62**	-.04	-.05	.31	-.03	.00
A2:	9	.09	-.21	**.70**	.23	-.16	-.02
	39	.00	.17	**.52**	.05	.16	.00
	69	.00	-.07	**.49**	.12	.18	.37
	99	.05	.15	.17	**.50**	-.13	.00
	129	-.09	.34	.15	.27	-.15	.16
	159	.00	.28	**.53**	.10	.06	-.01
	189	-.01	.36	**.46**	.01	.09	.01
	219	-.17	.00	**.59**	-.02	-.03	.21
A3:	14	-.10	**.50**	.07	-.20	.31	-.21
	44	.00	**.59**	.05	.13	-.04	.18
	74	-.06	**.62**	.26	.16	-.08	-.07
	104	-.06	**.62**	.11	.12	-.11	.20
	134	-.10	**.54**	-.19	.23	.00	-.02
	164	.03	**.75**	-.26	.00	.04	-.04
	194	.06	**.46**	-.07	.29	.00	-.12
	224	.09	.39	.00	**.55**	-.07	-.01
A4:	19	.06	.02	.24	.11	.21	.39
	49	.28	-.27	**.48**	-.09	.04	-.01
	79	.10	.01	.05	-.14	.05	**.64**
	109	.17	.28	.30	-.26	.00	**.44**
	139	.32	-.05	.00	.07	.04	.23
	169	.25	.19	.33	.03	-.01	.21

(table continues)

Table 3 (continues)

NEO PI-R Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
199	.33	.24	.20	-.17	.01	.00
229	.16	**.44**	.17	-.02	-.07	-.08
A5: 24	.09	-.06	-.02	-.04	**.60**	.21
54	.04	-.07	.00	.00	**.60**	.28
84	.03	.07	.30	.20	**.42**	-.13
114	.00	.39	-.11	-.09	**.41**	.23
144	-.28	-.12	.24	.04	**.56**	.06
174	.00	-.03	-.01	.08	**.61**	-.16
204	.32	.05	-.24	.19	.31	.18
234	-.05	-.06	.14	.19	**.52**	-.21
A6: 29	.05	.12	-.13	.23	.35	.01
59	.36	.12	.27	-.29	.07	-.08
89	-.14	.03	.07	**.48**	.26	-.01
119	.22	.02	.04	.06	.16	-.34
149	.04	.00	-.01	**.41**	.14	-.26
179	.24	.06	-.01	**.46**	.14	-.09
209	.05	.17	.01	**.55**	.12	-.08
239	-.02	-.11	.14	**.42**	.06	.27

Note: A1 = Trust; A2 = Straightforwardness; A3 = Altruism; A4 = Compliance; A5 = Modesty; A = 6 Tender-Mindedness. These are oblique promax-rotated facotrs - Procrustean transformation (standardized regression coeeficients). Loadings $\geq \pm$ .40 are given in boldface. $n$factor = 6.

however, to Feingold's Psych Bull meta-analysis on sex differences which examined a 26 country cross-cultural replication of the normative sex differences on the Agreeableness facets (Costa, Terracciano, & McCrae, 2001).

The results of the correlational analyses also confirmed the relationship between the facets, items, and domain of Agreeableness providing further support for this psychometric construct. All correlations were significant at the $p < .01$ level. While the intercorrelations between the Agreeableness facets were all significant, the most powerful relationship was with the Agreeableness domain.

Item Analysis

A confirmatory factor analysis was conducted using the maximum-likelihood method with an oblique model to exam the underlying factor structure of Agreeableness. Neither the proposed six factor model with 48 items nor the aggregated model with item parceling provided a satisfactory good fit to the data. All statistics of fit indices (e.g., GFI, AGFI, and RSMEA) of both models were unsatisfactory. In addition, the standardized estimates of parameter show that twelve out of forty-eight item loadings were below .40, among which six items were below .30. The results presented did not support the psychometrics of the items associated with the six facets of Agreeableness domain indicating that items did not appear to be valid indicators of the facets they are intended to measure which may be due to a lack of discriminant validity among the facets. While the facets and items reflect the domain of Agreeableness, some items did not reflect their intended facet.

Results of the exploratory factor analysis resulted in 35 of the 48 items significantly loading on their intended facet, suggesting that 13 items may be misplaced. While no multiple high loading items resulted, if the hypothesized structure of the items associated with the Agreeableness domain were supported, then all 48 items would have significantly loaded on their intended facet. Trust was the only facet with all eight items loading on their intended facet. The item that did not load on the facet of Altruism in factor 2 was (item 224: "I go out of my way to help others if I can") and with Modesty in factor 5 was (item 204: "I would rather praise others than be praised myself"). Next, factor 3, which is associated with the facet of Straightforwardness resulted in six of the eight intended items loading. Items that did not load on Straightforwardness include (items 99: "Being perfectly honest is a bad way to do business" and 129: "I would hate to be thought of as a hypocrite"). The three items that did not load on the facet of Tender-Mindedness include (items 29: "Political leaders need to be more aware of the human side of their policies"; 59: "I'm hard-headed and tough-minded in my attitudes"; 119: "I have no sympathy for panhandlers"). Finally, the facet of Compliance resulted in only two of the eight items loading on that facet. This facet, as well as the items that purportedly comprise this facet, appear to need further evaluation. Specifically, the items that did not load on Compliance should be further evaluated (items 19: "I would rather cooperate with others than compete with them"; 49: "I can be sarcastic and cutting when I need to be"; 139: "When I've been insulted, I just try to forgive and forget";

169: "If someone starts a fight, I'm ready to fight back"; 199: "I'm hard-headed and stubborn"; 229: "I often get into arguments with my family and co-workers").

While four items loaded on a different facet (items 49, 99, 224 and 229), the majority, nine items, did not load on any facet associated with Agreeableness. There may be a variety of explanations for this result. First, several of these items were related more to the external and not directly to the individual's own personality. For example, item 29: "Political leaders need to be more aware of the human side of their policies" seems external and possibly may lead to confounded ideas, including thoughts of liberalism and conservatism. Next, several items purportedly related to the facet of Compliance appear to be highly individualistic in nature; a characteristic that may not be the best fit for Compliance. For example, item 199: "I'm hard-headed and stubborn" is highly individualistic and should possibly be more overt and suggest a relationship between the individual and another person. Further, item 119: "I have no sympathy for panhandlers," intended to load on Tender-Mindedness, may elicit different connotations from individuals based on their personal experiences. Finally, while the items may be significantly related to their purported facet, the items that did not load seem to lack a powerful relationship. For example, while all the items that purport to load on the facet of Straightforwardness are significantly correlated ($p < .0001$) the two items that did not load (items 99 and 129) have the weakest, least powerful, relationship.

Conclusions

This study provided results of analyses of the facets and items associated with the Agreeableness domain of the NEO PI-R. An important influence to take into consideration when examining intercorrelations among items is Acquiescence bias; the tendency to agree with items regardless of content. Because single items are essentially unbalanced scales, they tend to correlate with all other items. This bias contributes to the formation of positively and negatively keyed factors when on theoretical grounds one would expect a single bipolar factor. This study illustrates the powerful distorting effects on an artifact operating on the level of items; indicating facet level analysis is the more appropriate method of analysis. Researchers examining assessment measures, whether they are personality assessments or educational assessments, need to be aware of the impact an item analysis may have when interpreting results.

As practitioners generally utilize an item review when exploring

incongruent results with their clients, these results also suggest caution in interpreting scores at the item level. Further research is suggested to determine why the items associated with some facets (e.g., Compliance) resulted in a weak factor structure when the facets themselves strongly correlate with the Agreeableness domain, confirming prior studies on the NEO PI-R. It is also quite possible that the item level is not designed to be interpreted. While a personality assessment was utilized for the purposes of this study, the implications are applicable to other types of assessments, such as educational assessments. Personality and educational researchers and users of any type of assessment need to be aware that performing item level factor analysis almost always produces misleading or un-interpretable results.

References

Botwin, M. (1995). Review of the Revised NEO Personality Inventory. In J. Conoley, & J. Impera (Eds.). *The 12ᵗʰ mental measurements yearbook* (pp. 862-863). Lincoln, Nebraska: Buros Institute of Mental Measurements.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press Inc.

Costa, P. T., Jr., & McCrae, R. R. (1992a). *The Revised NEO Personality Inventory.* Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P. T., Jr., & McCrae, R. R. (1992b). *NEO PI-R: Professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P. T., Jr., & McCrae, R. R. (1997). Stability and change in personality assessment: The Revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68,* 86-94.

Costa, P. T., Jr., McCrae, R.R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12,* 887-898.

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

Furnham, A. (1996). The big five verses the big four: The relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences, 21,* 303-307.

Juni, S. (1995). Review of the Revised NEO Personality Inventory. In J. Conoley, & J. Impera (Eds.). *The 12ᵗʰ mental measurements yearbook*

(pp. 863-868). Lincoln, Nebraska: Buros Institute of Mental Measurements.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9,* 151-173.

Loehlin, J., McCrae, R. R., Costa, P. T., Jr., & John, O. (1998). Heritabilities of common and measure-specific components of the Big Five personality factors. *Journal of Research in Personality, 32,* 431-453.

McCrae, R. R., & Costa, P. T., Jr. (1992). Discriminant validity of NEO-PI-R facet scales. *Educational and Psychological Measurement, 52,* 229-237.

McCrae, R. R., & John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60,* 175-215.

McCrae, R. R., Zonderman, A., Costa, P. T., Jr., Bond, M., & Paunonen, S. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70,* 552-566.

Piedmont, R. L., & Weinstein, H. P. (1993). A psychometric evaluation of the new NEO-PIR facet scales for agreeableness and conscientiousness. *Journal of Personality Assessment, 60*, 302-318.