Educational and Psychological Measurement
70(6) 1021–1041

DOI: 10.1177/0013164410378091
http://epm.sagepub.com
SAGE

An Item-Level Examination of the Factorial Validity of NEO Five-Factor Inventory Scores

Darrell M. Hull[1], A. Alexander Beaujean[2], Frank C. Worrell[3], and Aimee E. Verdisco[4]

Abstract

The NEO Five-Factor Inventory (NEO-FFI) is often used in field-based research and clinical studies as it is designed to measure the same personality dimensions as the longer NEO Personality Inventory in a shorter time frame. In this study, the authors examined the reliability and structural validity of the NEO-FFI scores at the item level in a large sample of Jamaican young adults (*N* = 1,021; ages 17-24 years). Across different factor estimation and rotation methods, many NEO-FFI items performed poorly. Likewise, the estimated reliability of the NEO-FFI scores was poor (except Conscientiousness) across different estimations of reliability. These items and scores were then compared with other studies of the NEO-FFI that reported item-level pattern/structure coefficients or reliability estimates. Similarities in item performance and low reliability estimates across studies suggest that the items, rather than cultural differences, account for much of the poor performance of the NEO-FFI scores, especially in the domains of Extraversion, Openness to Experience, and Agreeableness.

Keywords

adolescents, Jamaica, NEO-FFI, personality, validity

[1]University of North Texas, Denton, TX, USA
[2]Baylor University, Waco, TX, USA
[3]University of California, Berkeley, Berkeley, CA, USA
[4]Inter-American Development Bank, Washington, DC, USA

Corresponding Author:
Darrell M. Hull, Department of Educational Psychology, University of North Texas, 1155 Union Circle No. 311335, Denton, TX 76203-5017, USA
Email: darrell.hull@unt.edu

The Five-Factor Model (FFM) is currently one of the most widely recognized and accepted conceptualizations of personality (Digman, 1990; Hong, Paunonen, & Slade, 2008; John & Srivastava, 1999). According to the FFM, there are five basic personality traits: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). The five factors have demonstrated a relationship to many life outcomes, such as (a) the ability to cope with stress; (b) academic and vocational success and failure; (c) the establishment of romantic relationships, lifelong friendships, and life goals; and (d) health outcomes, such as mortality and accident incidence (Black, 2000; Block, 1993; Haan, Millsap, & Hartka, 1986; Helson & Moane, 1987; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007; Robins, Fraley, Roberts, & Trzesniewski, 2001). Moreover, there is consensus that individual differences in personality begin to exert their influence as early as adolescence; thus, they are important to measure when investigating the psychological processes in adolescents and emerging adults (Caspi, Roberts, & Shiner, 2005).

One of the more attractive qualities of the FFM has been its purported universality (McCrae, Terracciano, & Khoury, 2007). McCrae and Costa (1997) argued that the FFM is the best way to summarize personality both in English-speaking populations as well as cultures in Europe, the Middle East, and Asia. In 2002, McCrae and Allik devoted an entire book to the topic of the cross-cultural validity of the FFM, with the crux of most chapters indicating that although there may be mean and/or variability differences in personality across cultures, FFM traits are so universal that their existence might constitute a cross-cultural psychological law (cf. McCrae, 2002).

The NEO Personality Inventories

Much of the cross-cultural work done with the FFM has relied on the family of NEO Personality Inventories, such as the NEO Personality Inventory (NEO-PI; Costa & McCrae, 1989) and its successor, the NEO Personality Inventory–Revised (NEO-PI-R; Costa & McCrae, 1992). The NEO-PI and NEO-PI-R were designed to measure the five factors of the FFM using 30 different cluster scores (6 for each factor) derived from 180 items. After Costa and McCrae (1989) developed the NEO-PI, they developed a more parsimonious version of the instrument, selecting items from the NEO-PI via varimax rotation (McCrae & Costa, 1989) to maximize convergent and discriminant validity with the NEO-PI factors. The resulting instrument was the NEO Five-Factor Inventory (NEO-FFI), a 60-item instrument designed as a brief measure for each of the FFM factors. However, unlike the NEO-PI, the NEO-FFI did not undergo traditional development (i.e., norming and revision of items). Instead, using the results from the NEO-PI-R norming data, Costa and McCrae (1992) selected the 12 items from each of the five factors (60 items total) with the highest pattern/structure coefficients on the principal component that represented that particular personality domain (McCrae & Costa, 2004, 2007). The final 60 items that were included in the NEO-FFI were based on additional exploratory factor analyses and examinations of internal consistency.

The NEO family of instruments have been translated into multiple languages and used in multiple contexts with a variety of populations; in fact, Pytlik Zillig, Hemenover, and Dienstbier (2002) contended that the NEO instruments are some of the most widely used measures of the FFM. There have been multiple studies on the reliability and validity of NEO-FFI scores, as well as the cross-cultural transportability of the items (e.g., Caruso, 2000; Church et al., 2008; Holden, Wasylkiw, Starzyk, Book, & Edwards, 2006; Hong et al., 2008; McCrae & Costa, 2007; Murray, Rawlings, Allen, & Trinder, 2003; Scandell, 2000). In 2007, McCrae and Costa noted that the large number of citations (>680 as of December 2006) on the NEO-FFI in PsycINFO confirmed "the research utility of brief measures of the FFM" (p. 116).

Rank order stability of personality with the FFM suggests that cognitive ability is the only psychological construct that is more consistent (Conley, 1984; Fraley & Roberts, 2005; Roberts & DelVecchio, 2000). Despite life-changing roles and identity decisions in young adulthood that are greater in magnitude than any others during the life span, differences in personality are remarkably consistent throughout this period (Arnett, 2000). Sherry, Henson, and Lewis (2003), though, compared 79 adolescents and 80 college students using the NEO-PI-R norms and found differences on all five domains. However, coefficient alpha estimates for scores from the two groups were quite similar, with the adolescents' estimates ranging from .56 to .83 and the college students' estimates ranging from .66 to .84. Roth and Herzberg (2004) reported similar findings for the NEO-FFI in a sample of 1,236 German adolescents (M age = 14.89 years), and Hrebícková et al. (2002) also found similar coefficient alpha estimates for scores in samples of Polish (N = 350; M age = 16.6 years, SD = 1.1) and Slovak (N = 516; M age = 16.49 years, SD = 1.8) adolescents. Hrebícková et al. also reported that the five NEO-FFI factors in the adolescent groups were generally congruent with adult samples from Czechoslovakia, Germany, and the United States. In sum, the extant literature suggests that NEO-FFI scores have comparable reliability and structural validity in adolescent and adult samples.

Despite the widespread use of the NEO-FFI, there have been some criticisms of the reliability and structural validity of the instrument's scores (Becker, 2006; Egan, Deary, & Austin, 2000) and the psychometric procedures that Costa and McCrae (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996) advocated for examining the scores. In Caruso's (2000) reliability generalization study of all NEO instruments, he found that the mean reliability coefficients were .88, .83, .79, .75, and .83 for the N, E, O, A, and C personality domains, respectively. However, for the 20 NEO-FFI studies, the reliabilities were .83, .75, .65, .67, and .80 for these domains, suggesting that scores on the NEO-FFI's O and A subscales might not be measured well. This finding was replicated in a study by Egan et al. (2000), whose item-level analyses of the O scale in a large British sample showed that several items did not correlate strongly with the intended factor, irrespective of the factor analysis method.

Regarding the structural validity concern, McCrae and Costa (2004) and McCrae et al. (1996) argued for the use of two specific ways of analyzing NEO-PI-R and NEO-FFI data. They argued against confirmatory factor analysis (CFA) for structural

examinations of scores on the NEO instruments and, instead, recommended using a principal components estimation specifying five factors and then either rotating the components using a varimax rotation or a Procrustes rotation. As Procrustes rotation requires a target matrix to rotate toward, they specified two acceptable target matrices: either the hypothesized factor pattern/structure coefficients from the technical manual (i.e., a matrix consisting solely of 1s and 0s) or a pattern/structure coefficient matrix from a previously published study of the instrument being analyzed.

Although the psychometric methods advocated by McCrae and Costa (2004) might be justifiable for the NEO-PI-R because the indicator variables (i.e., the facet scores) are continuous—for another point of view, see Floyd and Widaman (1995) and Widaman (2007)—there is an abundance of literature indicating that it would not be good practice to analyze the NEO-FFI scores in this way because the indicator variables in this case are the items themselves (Morizot, Ainsworth, & Reise, 2007; Waller, Tellegen, McDonald, & Lykken, 1996). As neither principal component analysis (PCA) nor traditional exploratory factor analysis was designed for categorical indicators, factoring items without taking their categorical nature into account can result in multiple problems, including extracting extraneous factors (Bernstein & Teng, 1989). Recommended alternatives are to use either an exploratory factor model that accounts for these types of data or to use an item response theory (IRT) framework, as IRT analyses take the item difficulties into account (Bartholomew & Knott, 1999; McDonald, 1999; B. O. Muthén, Kao, & Burstein, 1991). Although a few researchers have conducted sound item-level analyses on the NEO-FFI (e.g., Huang, Church, & Katigbak, 1997; Reise, Smith, & Furr, 2001), they were concerned with specific factors and not the entire instrument.

Current Study

The purpose of the current study is threefold. The first goal is to examine the factor structure and reliability of NEO-FFI scores in a sample of young Jamaican adults. Although there have been investigations of the NEO instruments in many countries (e.g., McCrae, 2002) to date, little psychological research has been conducted in the Caribbean. The research that has been done (e.g., DeLaet & Wise, 1986; Hamilton & Leo-Rhynie, 1979; Hutchinson et al., 2004; Richardson, 1999a, 1999b) has used either instruments and classification systems validated in other countries, with the assumption that both the scores and norms are generalizable, or surveys and instruments developed by the researcher.

The second purpose of this study is to compare item-factor pattern/structure coefficients of the NEO-FFI using three different methods. To remove unneeded complexity, we subsume both PCA and IRT methods under the term *factor analysis*, although we are aware that principal component and factor analysis methods are not the same (Widaman, 2007). The first method is that recommended by McCrae et al. (1996), that is, a PCA of the items with a Procrustes rotation toward the theoretical target matrix (i.e., a matrix with 1s and 0s). The second and third methods are item-level analysis

techniques recommended by other methodologists (e.g., Bernstein & Teng, 1989; Panter, Swygert, Dahlstrom, & Tanaka, 1997): (a) item-level exploratory factor analysis (B. O. Muthén, 1979; B. O. Muthén & Kaplan, 1992) with orthogonal and oblique rotations and (b) item-level CFA (Flora & Curran, 2004). Although we did not expect to find substantial differences between the exploratory and confirmatory item-level analyses (Kay, 2004), using both types of analyses allowed for a more comprehensive examination of the structure of NEO-FFI items.

The third purpose of this study is to compare the results of the factor analysis of the Jamaican sample to other published works that included factor analyses of NEO-FFI items. These studies came from adult participants in Britain, Canada, France, Japan, Poland, the Slovak Republic, and the United States. Comparing the results from the current sample to these findings provided another index of the validity of NEO-FFI item scores in cross-cultural studies.

Method

Participants

Participants consisted of 1,021 youth (59.6% female, *M* age = 18.3 years; *SD* = 1.8) recruited by the Jamaican National Youth Service to participate in a longitudinal study on the efficacy of a youth development program. Jamaica is a small English-speaking country of approximately 2.8 million people, with 91% of the population being of African descent (Central Intelligence Agency, 2008). Assessing the psychometric properties of the NEO-FFI items and scores in a Jamaican sample represents a unique opportunity. As the native language in Jamaica is English, there is no apparent need to contend with the complications of translation of the instrument; yet the culture is different from that of the United States, and a large group of Jamaican respondents have not been used in previously published studies of the NEO-FFI.

The Jamaican youth population (i.e., ages 15-24 years) consists of more than 480,000 individuals, representing 18.3% of the country's total population (Planning Institute of Jamaica, 2006). The sample for the present study is representative of a group of young adults targeted for interventions in developing countries such as Jamaica referred to as *unattached*. Unattached youth are defined as those who do not have a job or are not currently enrolled in any educational or other training program. In Jamaica, conservative estimates put the number of unattached youth somewhere around 140,000 (Fox, 2004). These youth have higher rates of poverty and lower levels of education (Miller, 1999) than the general youth population. A convenience sample of the participants in three parishes (*N* = 79, 38% male) were individually administered a computer-adaptive reading assessment (Renaissance Learning, 2010). The mean grade equivalent score for this subsample was 6.3 (*SD* = 3.2). Costa and McCrae (1992) stated that the "NEO-FFI requires a sixth-grade reading level" (p. 4).

Instrument and Procedures

The NEO-FFI (Costa & McCrae, 1992) is a 60-item shortened version of the NEO-PI, which was developed to assess the domains of the five-factor model of personality: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (John & Srivastava, 1999). The NEO-FFI is frequently used in assessing the FFM, in part because of its brevity (it typically requires 15-20 minutes to complete; McCrae & Costa, 2007). The NEO-FFI requires participants rate themselves on a 5-point Likert-type scale (1 = *strongly disagree*; 5= *strongly agree*). The respondents completed the NEO-FFI along with a battery of other group-administered assessments. Project staff or graduate students from the University of the West Indies–Mona Campus supervised all administrations. As recommended by Costa and McCrae (1992), responses from four individuals who did not complete more than nine items were not included in analyses. Of the remaining 1,017 respondents, 11% *(n* = 110) were missing between one and nine items; consequently, all analyses were done using the full-information maximum-likelihood (FIML) estimator in Mplus (L. K. Muthén & Muthén, 2007), which allows for the use of incomplete data (Little & Rubin, 2002).

Data Analyses

The analysis steps for the Jamaican data were as follows. First, as recommended by McCrae and Costa (2004) and McCrae et al. (1996), all 60 NEO-FFI items were put into a PCA. This was done by first calculating the correlations among the items (using FIML), and then conducting the PCA on these correlations, extracting five components. Initially, the components were rotated using a varimax rotation, but then they were rotated using a Procrustes rotation using the vegan (Oksanen et al., 2009) package in R (R Development Core Team, 2006). The item pattern/structure coefficients were rotated toward the theoretical NEO-FFI target matrix of 1s and 0s (i.e., 1s for where an item should correlate and 0s for where an item should not load). Second, we conducted an item-level common factor analysis (IL-FA), which takes the item thresholds into account by factoring a matrix of polyserial correlations (using FIML), extracting five factors, and then rotating them using varimax rotation. In addition, we also rotated the IL-FA factor obliquely using the quartamin rotation in keeping with literature stating that the five factors are likely not orthogonal to each other (e.g., Blackman, Renwick, Donnelly & Logan, 2004; Borkenau & Ostendorf, 1990). Third, we did a CFA on each of the five factors in the NEO-FFI, constraining the latent variances to be unity; this constraint makes this analysis equivalent to an IRT analysis of the items (Kamata & Bauer, 2008).

In addition to conducting these analyses on the Jamaican data, we also examined item pattern/structure coefficients from other published NEO-FFI studies: Egan et al. (2000); Holden and Fekken (1994); Hrebícková et al. (2002); McCrae and Costa (2004); Parker and Stumpf (1998); Rolland, Parker, and Stumpf (1998); and Yoshimura, Ono, Nakamura, Nathan, and Suzuki (2001). These seven published studies included

Table 1. Summary of Samples From Studies That Reported Pattern Structure Coefficients on the NEO-FFI Used for Comparison With the Present Study

Authors	Sample *N*	Age (years)
Egan, Deary, and Austin (2000)	(Farmers) *N* = 252	Range = 25-83
	(Doctors) *N* = NA	
	(Mental patients) *N* = NA	
Holden and Fekken (1994)	*N* = 243	Median = 21.29
Hrebícková et al. (2002)	(Czech) *N* = 945	*M* = 24.34 (*SD* = 13.16)
	(Polish) *N* = 350	*M* = 16.6 (*SD* = 1.1)
	(Slovak) *N* = 516	*M* = 16.49 (*SD* = 1.8)
McCrae and Costa (2004)	*N* = 1,959	*M* = 16.5 (*SD* = 7)
	N = 1,492	*M* = 56.2 (*SD* = 17)
Parker and Stumpf (1998)	*N* = 870	*M* = 13.7 (*SD* = 0.96)
Rolland, Parker, and Stumpf (1998)	*N* = 447	*M* = 20.9 (*SD* = 3.2)
Yoshimura, Ono, Nakamura, Nathan, and Suzuki (2001)	(Males) *N* = 1,609	*M* = 43.4 (*SD* = 11.8)
	(Females) *N* = 1,729	*M* = 44.1 (*SD* = 12)

NEO-FFI = NEO Five-Factor Inventory; NA = not available.

11 factor analyses from adult participants in Japan, Britain, Canada, the United States, France, the Czech Republic, Poland, and the Slovak Republic, and one sample of precocious adolescents from the United States (Parker & Stumpf, 1998). Of the 13 samples in the 7 published studies, only 4 samples with reported age ranges (±1 *SD*) did not overlap with the ages of the Jamaican sample as shown in Table 1.

Finally, we estimated internal consistency estimates for each of the FFM personality domains using two methods: Cronbach's (1951) alpha and McDonald's (1999) omega. The former requires individual responses, so only people with complete responses within a domain could be used; consequently, there were different sample sizes for each personality domain, ranging from 977 (Openness to Experience) to 1,004 (Extraversion). The latter is calculated using item factor pattern/structure coefficients, and we used the coefficients from the CFA/IRT analysis.

Results

Table 2 shows the results from the multiple factor analyses of the NEO-FFI items. Across all data reduction techniques, there were similar patterns. Most items that had strong (or weak) pattern/structure coefficients on a factor from one estimation/rotation technique had coefficients of similar magnitudes from the other estimation/rotation techniques. Moreover, as the nonbolded pattern/structure coefficients indicate, there was no subscale in which all the intended items achieved salient pattern/structure coefficients. Correlations among subscales in the oblique rotation ranged from |.06| (between Extraversion and Agreeableness) to |.41| (between Extraversion and

Table 2. NEO Five-Factor Inventory Item-Level Pattern Coefficients

		PCA: Varimax					PCA: Procrustes					Item FA: Varimax					Item FA: Quartimin					
Item	Domain	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C	CFA/IRT
1	N	0.14	−0.04	0.07	0.02	0.02	0.12	0.00	0.07	−0.06	0.04	0.11	−0.01	0.03	−0.03	0.01	0.11	−0.02	0.04	−0.04	0.03	0.13
6	N	**0.34**	−0.01	−0.02	0.17	−0.05	**0.33**	0.05	−0.01	−0.18	−0.05	**0.30**	0.00	−0.03	−0.14	−0.05	0.29	0.00	−0.07	−0.11	0.03	**0.55**
11	N	**0.61**	0.04	−0.02	−0.11	0.07	**0.60**	0.08	0.09	0.08	0.11	**0.53**	0.01	0.00	0.08	0.05	**0.59**	−0.02	0.05	0.10	0.09	**0.50**
16	N	0.13	−0.28	0.05	0.12	−0.06	0.13	−0.21	0.09	−0.21	−0.03	0.12	−0.15	0.00	−0.12	−0.07	0.08	−0.18	0.02	−0.13	0.02	**0.49**
21	N	**0.51**	−0.07	0.12	0.06	−0.07	**0.48**	0.04	0.18	−0.15	−0.01	**0.45**	−0.05	0.10	−0.08	−0.08	**0.41**	−0.07	0.11	−0.14	−0.01	**0.47**
26	N	**0.52**	−0.06	−0.04	0.10	−0.19	**0.52**	0.00	0.08	−0.13	−0.17	**0.47**	−0.05	−0.03	−0.13	−0.19	**0.43**	−0.06	−0.02	−0.11	−0.09	−0.18
31	N	0.04	−0.11	0.00	−0.15	−0.11	0.05	−0.13	0.11	0.10	−0.08	0.03	−0.07	−0.03	0.08	−0.10	0.03	−0.10	0.03	0.07	−0.07	−0.06
36	N	**0.63**	0.00	−0.05	0.07	−0.01	**0.63**	0.07	0.03	−0.09	0.01	**0.57**	0.00	−0.02	−0.09	−0.03	**0.57**	−0.02	−0.02	−0.07	0.07	0.18
41	N	**0.69**	0.03	−0.13	−0.04	−0.14	**0.70**	0.06	0.04	0.03	−0.12	**0.66**	0.01	−0.11	0.02	−0.14	**0.68**	0.00	−0.07	0.09	−0.07	0.25
46	N	**−0.33**	0.06	−0.03	−0.02	0.15	**−0.33**	0.01	−0.13	0.07	0.11	−0.28	0.06	−0.02	0.06	0.14	−0.25	0.06	−0.04	0.06	0.10	**0.46**
51	N	**0.49**	0.06	−0.01	0.29	−0.21	**0.48**	0.15	0.01	−0.28	−0.21	**0.46**	0.06	0.00	−0.27	−0.19	**0.39**	0.06	−0.05	−0.25	−0.12	0.09
56	N	**0.57**	0.05	0.03	0.00	−0.09	**0.55**	0.11	0.13	−0.03	−0.04	**0.51**	0.03	0.03	−0.04	−0.09	**0.51**	0.01	0.06	−0.03	−0.05	**0.61**
2	E	−0.08	**0.50**	0.13	0.18	−0.01	−0.13	**0.53**	−0.06	−0.05	−0.04	−0.04	**0.39**	0.10	−0.08	0.04	−0.03	**0.45**	0.00	−0.08	−0.10	**0.37**
7	E	0.22	**0.39**	0.08	−0.10	−0.09	0.18	**0.39**	0.09	0.15	−0.08	0.18	0.24	0.09	0.07	−0.06	0.19	0.26	0.08	0.05	−0.16	0.10
12	E	0.00	0.15	−0.01	−0.04	0.03	0.00	0.13	−0.03	0.09	0.02	−0.01	0.12	0.00	0.04	0.03	0.02	0.11	0.00	0.06	−0.01	**0.55**
17	E	0.01	**0.57**	0.07	0.01	0.27	−0.04	**0.56**	−0.12	0.13	0.23	0.01	**0.48**	0.08	0.07	0.27	0.09	**0.49**	0.01	0.08	0.10	0.13
22	E	0.09	**0.42**	0.08	0.18	−0.12	0.05	**0.46**	−0.02	−0.08	−0.15	0.10	0.29	0.08	−0.11	−0.07	0.08	**0.32**	0.01	−0.12	−0.16	0.11
27	E	**−0.35**	**0.33**	−0.03	−0.17	−0.14	**−0.35**	0.23	−0.04	0.28	−0.18	−0.29	0.25	−0.05	0.18	−0.10	−0.25	0.28	−0.04	0.21	−0.24	**0.46**
32	E	−0.06	0.07	**0.36**	0.09	0.08	−0.13	0.18	0.22	−0.17	0.15	−0.05	0.08	0.26	−0.05	0.10	−0.06	0.11	0.20	−0.15	0.04	**−0.31**
37	E	−0.17	**0.48**	0.20	0.11	**0.38**	−0.24	**0.52**	−0.10	−0.01	**0.35**	−0.16	**0.48**	0.16	−0.03	**0.38**	−0.08	**0.52**	0.05	−0.04	0.19	**0.41**
42	E	−0.24	**0.37**	0.20	−0.18	0.15	−0.29	**0.35**	0.09	0.22	0.17	−0.22	**0.31**	0.16	0.20	0.16	−0.15	**0.34**	0.15	0.15	−0.03	−0.10
47	E	0.04	0.19	0.19	0.27	−0.02	−0.02	0.28	0.04	−0.26	−0.01	0.06	0.15	0.12	−0.18	0.02	0.03	0.17	0.05	−0.21	−0.02	0.15
52	E	−0.17	**0.33**	0.13	0.28	**0.40**	−0.22	**0.39**	−0.20	−0.19	**0.35**	−0.15	**0.34**	0.11	−0.18	**0.40**	−0.10	**0.38**	−0.04	−0.17	**0.30**	**0.53**

(continued)

Table 2. (continued)

		PCA: Varimax					PCA: Procrustes					Item FA: Varimax					Item FA: Quartimin					
Item	Domain	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C	CFA/IRT
57	E	**–0.31**	0.07	0.24	–0.24	0.19	**–0.36**	0.06	0.17	0.19	0.24	–0.29	0.08	0.18	0.24	0.19	–0.22	0.09	0.20	0.17	0.06	**0.44**
3	O	0.16	0.03	0.07	0.23	–0.26	0.14	0.10	0.07	–0.24	–0.24	0.16	0.04	0.03	–0.20	–0.21	0.09	0.04	0.00	–0.21	–0.19	**0.51**
8	O	0.03	–0.07	0.04	0.12	**–0.46**	0.03	–0.04	0.14	–0.15	**–0.43**	0.05	–0.04	–0.01	–0.14	**–0.40**	–0.06	–0.02	0.01	–0.15	**–0.39**	**0.55**
13	O	0.08	–0.02	**0.53**	–0.27	0.03	–0.02	0.09	**0.56**	0.07	0.20	0.05	–0.02	**0.45**	0.22	0.03	0.05	–0.01	**0.50**	0.00	–0.06	**0.38**
18	O	–0.03	0.05	0.03	–0.26	0.04	–0.03	0.00	0.09	0.25	0.07	–0.05	0.03	0.02	0.19	0.04	0.00	0.03	0.08	0.18	–0.02	0.26
23	O	–0.10	0.00	**0.48**	**–0.32**	0.07	–0.19	0.07	**0.49**	0.14	0.22	–0.11	0.03	**0.38**	0.26	0.08	–0.08	0.00	**0.50**	0.07	–0.06	**0.54**
28	O	0.04	0.01	**0.47**	0.17	0.00	–0.06	0.18	**0.34**	**–0.30**	0.11	0.05	0.05	**0.33**	–0.12	0.04	0.00	0.07	0.28	–0.26	–0.01	0.09
33	O	0.02	–0.05	0.01	–0.16	–0.23	0.03	–0.08	0.15	0.12	–0.20	0.02	–0.05	–0.01	0.08	–0.20	0.00	–0.06	0.05	0.08	–0.21	**0.58**
38	O	–0.07	–0.06	**–0.32**	–0.05	–0.19	0.01	–0.17	–0.20	0.13	–0.25	–0.05	–0.08	–0.25	0.02	–0.17	–0.05	–0.07	–0.23	0.12	–0.13	**0.48**
43	O	0.03	0.03	**0.58**	0.02	0.23	–0.09	0.20	**0.41**	–0.19	**0.36**	0.02	0.08	**0.49**	0.00	0.22	0.01	0.05	**0.51**	–0.22	0.13	**0.45**
48	O	–0.02	0.01	**0.34**	**–0.50**	0.04	–0.08	0.01	**0.46**	**0.35**	0.18	–0.05	–0.02	0.29	**0.41**	0.03	0.00	–0.04	**0.41**	0.26	–0.09	**0.34**
53	O	–0.02	0.11	**0.44**	–0.06	**0.31**	–0.12	0.22	0.28	–0.04	**0.41**	–0.04	0.12	**0.37**	0.08	**0.30**	0.00	0.13	**0.34**	–0.06	0.19	**0.54**
58	O	0.11	0.07	**0.35**	0.09	–0.04	0.04	0.19	0.28	–0.18	0.05	0.10	0.06	0.25	–0.06	–0.01	0.07	0.08	0.21	–0.17	–0.04	**0.70**
4	A	0.04	**0.40**	0.07	**–0.30**	**0.33**	0.00	**0.35**	0.01	**0.36**	**0.33**	0.02	**0.31**	0.09	**0.31**	**0.30**	0.15	**0.31**	0.11	**0.30**	0.15	0.20
9	A	**–0.36**	0.04	–0.12	–0.25	**0.35**	**–0.33**	–0.08	–0.18	**0.31**	**0.30**	**–0.34**	0.02	–0.10	0.26	**0.31**	–0.23	0.00	–0.06	**0.30**	0.26	**0.66**
14	A	**–0.43**	0.09	–0.04	–0.22	0.12	**–0.42**	–0.01	–0.07	0.27	0.09	**–0.38**	0.06	–0.04	0.23	0.11	**–0.32**	0.05	–0.01	0.24	0.03	0.09
19	A	0.07	**0.33**	–0.08	–0.21	0.24	0.07	0.26	–0.11	**0.31**	0.20	0.03	0.21	–0.01	0.19	0.20	0.12	0.21	–0.01	0.22	0.12	0.21
24	A	–0.20	0.11	0.04	–0.26	–0.07	–0.20	0.04	0.10	0.27	–0.05	–0.17	0.06	0.02	0.20	–0.04	–0.15	0.06	0.08	0.19	–0.14	**0.32**
29	A	–0.20	0.00	–0.01	0.09	**–0.32**	–0.19	–0.01	0.02	–0.07	**–0.32**	–0.13	–0.01	–0.04	–0.07	–0.24	–0.18	0.01	–0.04	–0.06	–0.25	**0.50**
34	A	–0.16	**0.37**	0.04	–0.10	0.29	–0.19	**0.33**	–0.11	0.20	0.26	–0.15	0.28	0.05	0.15	0.28	–0.06	0.28	0.02	0.15	0.16	**0.34**
39	A	–0.20	0.16	–0.04	**–0.51**	0.10	–0.18	0.02	0.07	**0.54**	0.11	–0.20	0.07	–0.02	**0.44**	0.08	–0.11	0.06	0.09	**0.43**	–0.05	**0.45**
44	A	–0.21	0.12	–0.09	**–0.38**	0.25	–0.19	0.00	–0.06	**0.43**	0.23	–0.21	0.07	–0.06	**0.35**	0.21	–0.10	0.06	0.01	**0.37**	0.12	**0.47**

(continued)

Table 2. (continued)

		PCA: Varimax					PCA: Procrustes					Item FA: Varimax					Item FA: Quartimin					
Item	Domain	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C	CFA/IRT
49	A	–0.03	0.18	0.15	–0.22	**0.53**	–0.08	0.18	0.01	0.21	**0.55**	–0.05	0.14	0.16	0.24	**0.49**	0.08	0.12	0.16	0.19	**0.40**	**0.50**
54	A	–0.17	0.03	0.03	**–0.37**	–0.25	–0.16	–0.06	0.21	**0.34**	–0.21	–0.15	–0.01	0.00	0.28	–0.21	–0.14	–0.01	0.11	0.25	–0.29	**0.56**
59	A	–0.18	0.02	–0.13	**–0.39**	0.13	–0.14	–0.11	–0.03	**0.41**	0.12	–0.18	–0.01	–0.09	**0.31**	0.10	–0.09	–0.04	–0.01	**0.34**	0.06	**0.52**
5	C	–0.04	0.03	0.02	–0.04	**0.58**	–0.06	0.03	–0.15	0.05	**0.56**	–0.06	0.04	0.04	0.09	**0.52**	0.06	0.02	0.00	0.10	**0.51**	**0.44**
10	C	–0.08	0.06	0.10	–0.01	**0.56**	–0.11	0.08	–0.10	0.01	**0.55**	–0.09	0.05	0.11	0.06	**0.51**	0.02	0.02	0.06	0.03	**0.52**	**–0.32**
15	C	–0.17	–0.01	0.22	–0.11	–0.02	–0.20	0.01	0.21	0.04	0.04	–0.14	0.02	0.14	0.09	0.01	–0.14	0.02	0.16	0.03	–0.07	0.12
20	C	–0.05	0.12	0.25	–0.20	**0.52**	–0.11	0.15	0.11	0.15	**0.57**	–0.07	0.09	0.23	0.24	**0.48**	0.06	0.08	0.23	0.16	**0.39**	**0.68**
25	C	–0.22	0.19	0.16	0.13	**0.49**	–0.27	0.24	–0.12	–0.09	**0.46**	–0.20	0.18	0.14	–0.05	**0.47**	–0.13	0.19	0.04	–0.07	**0.41**	–0.10
30	C	**–0.32**	–0.03	0.02	–0.26	**0.51**	**–0.32**	–0.09	–0.08	0.25	**0.50**	**–0.31**	–0.03	0.02	0.28	**0.47**	–0.18	–0.06	0.05	0.28	**0.42**	0.23
35	C	–0.15	0.08	0.09	0.11	**0.72**	–0.18	0.12	–0.22	–0.08	**0.68**	–0.15	0.08	0.08	–0.06	**0.71**	–0.04	0.07	–0.03	–0.07	**0.71**	**0.43**
40	C	–0.18	–0.01	0.25	–0.10	**0.55**	–0.24	0.03	0.06	0.03	**0.59**	–0.19	–0.01	0.23	0.13	**0.51**	–0.10	–0.04	0.21	0.03	**0.49**	**0.32**
45	C	**–0.48**	–0.09	0.10	–0.12	0.29	**–0.49**	–0.12	–0.01	0.09	**0.30**	**–0.44**	–0.07	0.08	0.13	0.28	**–0.38**	–0.08	0.08	0.09	0.25	**0.44**
50	C	–0.21	0.06	0.12	0.16	**0.68**	–0.25	0.11	–0.20	–0.13	**0.65**	–0.20	0.08	0.10	–0.09	**0.68**	–0.10	0.05	–0.02	–0.10	**0.69**	**0.55**
55	C	–0.27	–0.03	0.06	–0.25	**0.48**	–0.28	–0.08	–0.03	0.23	**0.49**	–0.27	–0.03	0.06	0.27	**0.44**	–0.15	–0.05	0.08	0.25	**0.39**	**0.71**
60	C	–0.08	0.08	0.10	0.01	**0.73**	–0.11	0.12	–0.16	0.00	**0.71**	–0.08	0.08	0.10	0.05	**0.70**	0.06	0.06	0.02	0.04	**0.69**	**0.69**

N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; PCA = principal component analysis; Item FA = item-level factor analysis; CFA/IRT = confirmatory factor analysis/item response theory. Pattern/structure coefficients ≥ |.30| are in boldface. Factor pattern and structure coefficients are the same for all rotations except quartimin. The structure coefficients for the quartimin rotation are not shown because of space considerations but were generally close in magnitude and direction to those of the pattern coefficients.

Conscientiousness), with a median value of |.23|. The table of correlations can be obtained from the authors.

To judge if an item performed well, we used the following criteria recommended by Kline (2000): (a) an item had a pattern/structure coefficient ≥ |.30| for its intended (home) factor; (b) the coefficient's direction on the home factor was the same as that for a majority of the other 11 items intended to measure the same factor; and (c) the item either had no coefficients either ≥ *.30 or ≤ *.30 (where * is the coefficient's direction for the home factor) for any of the other four factors. For the CFA/IRT analysis, all pattern/structure coefficients on nonhome factors were constrained to be zero. The percentage of items that performed well across the estimation techniques are given in Table 3.

As the results in Table 3 indicate, a majority of items on the N and C subscales performed well across estimation and rotation methods. Results were less positive for the other subscales. Averaging across all five estimation/rotation techniques, the percentage of items that performed well for the domains were as follows: N (62%), E (41%), O (47%), A (45%), and C (82%). These results indicate that only one factor—Conscientiousness—consistently had more than 67% of their items performing well; three of the other four factors had, on average, less than 50% of items performing well. In sum, 73% of the 60 NEO-FFI items did not perform well in at least one estimation/rotation.

To check if the large number of items not performing well was specific to the current sample, we also examined how well items performed in 11 factor matrices from 7 separate published studies. To determine if an item performed well, we used the same criteria as we did with the item for the Jamaican sample. Table 4 has these results and shows that 72% of the items that did not perform well in the Jamaican sample did not perform well in the other factor analytic studies. When the poorly performing items are categorized by personality domain, the percentage of items are N (8%), E (50%), O (50%), A (50%), and C (0%).

The reliability estimates for scores on each NEO-FFI domain are provided in Table 5. Other than for Conscientiousness and Agreeableness, reliability estimates were lower than those typically reported for NEO-FFI scores (Caruso, 2000), and except for Conscientiousness, the estimates were lower than typically found for sound personality questionnaires (Kline, 1998; see Henson, 2001, for a discussion of the importance of reliability estimates). This pattern was the same for both the alpha (Cronbach, 1951) and omega (McDonald, 1999) reliability indices.

Discussion

In this study, we examined the structural validity and reliability of NEO-FFI scores in a sample of Jamaican young adults, and compared the results with the extant literature examining the properties of the NEO-FFI items and domain scores. Using multiple forms of factor analyses, the findings indicated that there were items in each of the five NEO-FFI personality domains that functioned well in the current sample.

Table 3. Percentage of Items Performing Well on the NEO Five-Factor Inventory for Across All Factor Estimation and Rotation Methods

Estimation	Rotation	N	E	O	A	C	All items
Principal component analysis	Varimax	67	50	50	42	83	58
Principal component analysis	Procrustes	67	42	25	42	92	53
Item-level common factor analysis	Varimax	67	25	33	25	83	47
Item-level common factor analysis	Quartimin	58	42	42	42	83	53
Confirmatory factor analysis/item response theory		50	46	83	75	67	65
All estimation/rotation techniques		62	41	47	45	82	55

N = Neuroticism; E = Extraversion; O = Openness C = Conscientiousness. Confirmatory factor analysis/item response theory analyses were done within each personality domain.

However, only two personality domains (N and C) had a majority of the items that performed well. To examine if the poor performance of the items was related to the sample, we examined how well the poor-performing items performed in 11 other published analyses of NEO-FFI scores that included item-level pattern/structure coefficients. The results indicated that the majority of the items that performed poorly in the Jamaican sample also performed poorly in other published analyses of the items. Thus, these findings indicate that despite the popularity of the NEO-FFI scale, there are structural validity concerns with regard to the scores.

With the exception of C, psychometric problems were also evident in an analysis of the internal consistency estimates for NEO-FFI scores. Although it was not entirely surprising that the omega coefficients were low, given the problems with the structural validity, the low alpha coefficients, especially for the E, A, and O scores raise concerns about the integrity of these subscales (Kline, 1998), especially in light of the average reliability estimates found for these scores in previous studies (Caruso, 2000).

Revisiting the Five Factors

As noted above, the psychometric properties of NEO-FFI items and scores found in this study are not uncommon, as there are several studies that have suggested that NEO-FFI scores may not be robust. For example, using a large sample from Britain, Egan et al. (2000) found problems with 14 NEO-FFI items, using both exploratory and confirmatory analyses, and reported that the items from the O and E domains did not produce very clean factors. Using American undergraduate students, Becker (2006) reported problems with the NEO-FFI's E and O factors and dissimilarity between sexes in the A and C domains. Hong et al. (2008) had to drop the O domain completely to obtain an acceptable model fit for NEO-FFI scores. Finally, the NEO-FFI authors (Costa & McCrae, 1992), themselves, identified 14 problematic items on the NEO-FFI, although only 8 of 25 poorly performing items in the current study (3, 8, 9, 24, 29, 34, 38, and 49) overlapped with items identified by McCrae and Costa (2004). A closer examination of the items reveals patterns that may be useful in revising the scale.

Table 4. Poorly Performing NEO Five-Factor Inventory Items

Item	Domain	Jamaican sample (5 analyses)	Studies in literature (11 factor matrices)	All studies (15 analyses)
46	N	4 (80)	4 (36)	8 (50)
27	E	3 (60)	6 (55)	9 (56)
32	E	5 (100)	4 (36)	9 (56)
47	E	5 (100)	4 (36)	9 (56)
52	E	4 (80)	6 (55)	10 (63)
57	E	4 (80)	6 (55)	10 (63)
3	O	4 (80)	8 (73)	12 (75)
8	O	4 (80)	10 (91)	14 (88)
18	O	5 (100)	5 (45)	10 (63)
28	O	2 (40)	4 (36)	6 (38)
33	O	4 (80)	5 (45)	9 (56)
38	O	4 (80)	10 (91)	14 (88)
4	A	3 (60)	6 (55)	9 (56)
9	A	3 (60)	5 (45)	8 (50)
19	A	4 (80)	4 (36)	8 (50)
29	A	4 (80)	6 (55)	10 (63)
34	A	4 (80)	9 (82)	13 (81)
49	A	4 (80)	9 (82)	13 (81)

N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness. Items shown had problems in more than 30% of the analyses in both the current sample and samples from the literature. Poor performance was defined as not having a pattern/structure coefficient of at least .30 on home factor and/or coefficients greater than .30 on nonhome factor. Numbers in parentheses are percentages.

Table 5. NEO Five-Factor Inventory (Neo-FFI) Reliability Estimates

Domain	Cronbach's α	McDonald's ω	NEO-FFI reliability generalization
Neuroticism	.65	.60	.83
Extraversion	.58	.47	.75
Openness to Experience	.30	.29	.65
Agreeableness	.62	.53	.67
Conscientiousness	.83	.72	.80

Reliability generalization coefficients are from Caruso (2000, p. 246).

Neuroticism is one of three factors (including Openness and Conscientiousness) that Rolland (2002) identified as showing "cross cultural generalizability" (p. 7). Eight of the 12 items on N worked well in the current study, and the majority of coefficients

for seven of these were greater than .45. Interestingly, the four items that did not work well (1, 16, 31, and 46) were all negatively worded items. Although the inclusion of reverse-coded items is a typical strategy for reducing response sets, there is also evidence that these items can pose problems in structural analyses (Weems & Onwuegbuzie, 2001). The fact that only one of the four items did not work well in the comparison studies suggests that the current sample may have had particular difficulties in responding to these negatively worded items.

Five of the 12 Extraversion items did not perform well in the current study or in the comparison studies. However, this result is misleading, as E is one of the factors that has shown little cross-cultural generalizability. A closer examination of the current results indicates that only three of the E items (2, 17, 37) had three or more of the five coefficients in the .48 and greater range. The other nine items had a majority of coefficients less than .40, with five of the nine having a majority of coefficients less than .30. Interestingly, the three items with the highest coefficients seem to better reflect the *social* nature of the E (i.e., "I like having people around, enjoy talking to people, am a cheerful, high-spirited person") than the other items with very low coefficients (e.g., "life is fast-paced, I'm bursting with energy").

Openness to Experience is another factor that has received some support for cross-cultural generalizability (Rolland, 2002). However, 6 of the 12 O items (i.e., 50%) had problems both in the current study and in the comparison studies from the literature. Moreover, an examination of the coefficients indicates that only three of the O items had substantial coefficients with median values of .49 or higher. These three items all mentioned poetry or art, suggesting that the core factor in this study was more a reflection of a love of aesthetic experiences rather than a reflection of a wide range of interests. Nine of the items had coefficients that were generally low to very low, including items on speculating about the nature of the universe (median coefficient = .34), having a lot of intellectual curiosity (median coefficient = .37), and playing with abstract theories and ideas (median coefficient = .28).

Agreeableness is a factor that has not generalized well across cultural contexts (Rolland, 2002). In the current study, 6 of the 12 items on this factor did not work well in both the current study and in the comparison studies. However, 11 of the 12 items had median coefficients lower than |.40| across the five different analyses, indicating that very few of the items had substantial correlations with the factor.

Conscientiousness was the only factor in which the majority of the items worked well in both the current study and the comparison studies. In all, 10 of the 12 items had substantial coefficients across the five analytic strategies, with median coefficients greater than .45. The two items with low coefficients (15 and 45) were both negatively worded items, and they worked well in the comparison studies, despite not working well in the current study.

It is also worth mentioning that recent studies (see, e.g., Deyoung, 2006; Erdle, Irwing, Rushton & Park, 2010; Musek, 2007; and Rushton & Irwing, 2008) have provided support for Digman's (1997) report of two higher order factors of personality. A Stability factor consists of Agreeableness, Conscientiousness, and an inverse of the

Table 6. Factor Intercorrelations From Quartimin Rotation

	Personality domain			
	N	E	O	A
E	–0.12			
O	–0.10	0.25		
A	–0.28	0.06	0.20	
C	–0.29	0.41	0.33	0.18

N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness.

Neuroticism scale referred to as Emotional Stability, and a Plasticity factor consists of Extraversion and Openness. The presence of higher order factors might explain the lack of consistency and generalizability in the five-factor model; however, only two of the factor intercorrelations from the quartimin rotation were greater than .30 in this study (see Table 6). Nevertheless, the potential presence of higher order factors was not explicitly explored in this study as our purpose was primarily to mirror recommended solutions by the test authors.

Limitations and Conclusion

This study has several limitations that center on the sample. The first is that NEO-FFI scores had not been used in this country before. Despite the claims of cross-cultural generalizability of the instrument's scores, it is possible that variables specific to this context (e.g., local idiomatic differences) could have influenced the findings. Second, the sample consisted of adolescents, and although the NEO-FFI has been examined in other adolescent samples, the instrument was developed using older samples. A third concern is the sample's education level. The adolescents in this sample are not college bound; indeed, they applied to participate in this program to be prepared for a vocation as they do not have the achievement levels necessary to attend college. The impact of this variable on NEO-FFI scores is also unknown.

Limitations notwithstanding, the results from this study indicate that the NEO-FFI scores are measuring C, and to some extent N, well in the Jamaican sample, and more important, these results are similar to findings from other studies examining NEO-FFI items. However, as at least half of the items in the other three domains are not performing well and the scores for these domains have both low reliability and poor structural validity, the results raise questions about the viability of the NEO-FFI as an operationalization of the Big Five. These findings have several implications. The first is that the instrument should be revised and some suggestions have been provided with regard to four domains. The revision process should specifically include cross-cultural examinations of items, as one of the authors' stated goals is to measure the FFM across

cultures and nationalities. Another concern centers on the use of NEO-FFI scores in making high-stakes decisions. The NEO-FFI is currently used in clinical and vocational settings and in a large number of countries, but the results of this study suggest that these uses for the scores are premature. Given the alignment of the poorly functioning items in the current Jamaican sample with samples from other studies representing several different countries, the results suggest that the source of the problem is with the items and not the samples or the cultural contexts. At best, the results indicate that there may be some scale development work needed to make NEO-FFI scores reliable and structurally valid as an appropriate brief measure for use in assessing the five-factor model of personality.

Authors' Note

The views and opinions expressed in this publication are those of the authors and do not necessarily reflect the official position of the Inter-American Development Bank.

Acknowledgments

We would like to express our gratitude to L. O. Williams, J. K. Tindigarukayo, and D. Fearon, who managed and supervised the data collection and administration of project staff at the Social Welfare Training Centre of the Open Campus, The University of the West Indies.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

This article was written with financial support from the Inter-American Development Bank.

References

Arnett, J. J. (2000). Emerging adulthood. A theory of development from the late teens through the twenties. *American Psychologist, 55,* 469-480. doi:10.1037/0003-066X.55.5.469

Bartholomew, D., & Knott, M. (1999). *Kendall's library of statistics 7: Latent variable models and factor analysis*. London, England: Edward Arnold.

Becker, G. (2006). NEO-FFI scores in college men and women: A view from McDonald's unified treatment of test theory. *Journal of Research in Personality, 40,* 911-941. doi:10.1016/j.jrp.2005.09.009

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Evidence for multidimensionality due to item categorization. *Psychological Bulletin, 105,* 465-477. doi:10.1037/0033-2909.105.3.467

Black, J. (2000). Personality testing and police selection: Utility of the "Big Five." *New Zealand Journal of Psychology, 29,* 2-9.

Blackman, R., Renwick, S. J. D., Donnelly, J. P., & Logan, C. (2004). Big five or big two? Superordinate factors in the NEO Five Factor Inventory and the Antisocial Personality Questionnaire. *Personality and Individual Differences, 37,* 957-970. doi:10.1016/j.paid.2003.10.017

Block, J. (1993). Studying personality the long way. In D. C. Funder, R. D. Parke, C. Tomlinson Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 9-41). Washington, DC: American Psychological Association. doi:10.1037/10127-018

Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences, 11,* 515-524. doi:10.1016/0191-8869(90)90065-Y

Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement, 60,* 236-254. doi:10.1177/00131640021970484

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56,* 453-484. doi:10.1146/annurev.psych.55.090902.141913

Central Intelligence Agency. (2008). *World factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/

Church, A. T., Katigbak, M. S., Reyes, J. A. S., Salanga, M. G. C., Miramontes, L. A., & Adams, N. B. (2008). Prediction and cross-situational consistency of daily behavior across cultures: Testing trait and cultural psychology perspectives. *Journal of Research in Personality, 42*, 1199-1215. doi:10.1016/j.jrp.2008.03.007

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality and self-opinion. *Personality and Individual Differences, 5,* 11-25. doi:10.1016/0191-8869(84)90133-8

Costa, P. T., Jr., & McCrae, R. R. (1989). *The NEO-PI/NEO-FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334. doi:10.1007/BF02310555

DeLaet, T. J., & Wise, S. L. (1986). The development and validation of the Personality Priorities Inventory for Adolescents. *Educational and Psychological Measurement, 46,* 455-459. doi:10.1177/001316448604600222

Deyoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91,* 1138-1151. doi:10.1037/0022-3514.91.6.1138

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Reviews of Psychology, 41,* 417-440.

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

Egan, V., Deary, I., & Austin, E. (2000). The NEO-FFI: Emerging British norms and an item-level analysis suggest N, A, and C are more reliable than O and E. *Personality and Individual Differences, 29,* 907-920. doi:10.1016/S0191-8869(99)00242-1

Erdle, S., Irwing, P., Rushton, J. P., & Park, J. (2010). The general factor of personality and its relation to self-esteem in 628,640 Internet respondents. *Personality and Individual Differences, 48,* 343-346. doi:10.1016/j.paid.2009.09.004

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466-491. doi:10.1037/1082-989X.9.4.466

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286-299. doi:10.1037/1040-3590.7.3.286

Fox, K. (2004). *Jamaica youth activity survey, 2002*. Kingston, Jamaica: The Statistical Institute of Jamaica.

Fraley, C., & Roberts B. W. (2005). Patterns of continuity: A dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course. *Psychological Review, 112,* 60-74. doi:10.1037/0033-295X.112.1.60

Haan, N., Millsap, R., & Hartka, E. (1986). As time goes by: Change and stability in personality over fifty years. *Psychology and Aging, 1,* 220-232. doi:10.1037/0882-7974.1.3.220

Hamilton, M., & Leo-Rhynie, E. (1979). Sex-role stereotyping and education: The Jamaican perspective. *Interchange, 10*(2), 46-56. doi:10.1007/BF01191439

Helson, R., & Moane, G. (1987). Personality change in women from college to midlife. *Journal of Personality and Social Psychology, 53,* 176-186. doi:10.1037/0022-3514.53.1.176

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34,* 177-189.

Holden, R. R., & Fekken, G. C. (1994). The NEO Five-Factor Inventory in a Canadian context: Psychometric properties for a sample of university women. *Personality and Individual Differences, 17,* 441-444. doi:10.1016/0191-8869(94)90291-7

Holden, R. R., Wasylkiw, L., Starzyk, K. B., Book, A. S., & Edwards, M. J. (2006). Inferential structure of the NEO Five-Factor Inventory: Construct validity of the big four personality clusters. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement,* 38, 24-40.

Hong, R. Y., Paunonen, S. V., & Slade, H. P. (2008). Big five personality factors and the prediction of behavior: A multitrait-multimethod approach. *Personality and Individual Differences, 45,* 160-166. doi:10.1016/j.paid.2008.03.015

Hrebícková, M., Urbánek, T., Cermák, I., Szarota, P., Ficková, E., & Orlická L. (2002). The NEO-Five Factor Inventory in Czech, Polish, and Slovak contexts. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 53-78). New York, NY: Kluwer Academic/Plenum.

Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. *Journal of Cross-Cultural Psychology, 28,* 192. doi:10.1177/0022022197282004

Hutchinson, G., Simeon, D. T., Bain, B. C., Wyatt, G. E., Tucker, M. B., & Lefranc, E. (2004). Social and health determinants of well being and life satisfaction in Jamaica. *International Journal of Social Psychiatry, 50,* 43-53.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York, NY: Guilford Press.

Kamata, A., & Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling, 15,* 136-153.

Kay, C. A. (2004). *A comparison of traditional and IRT factor analysis*. Unpublished doctoral dissertation, University of North Texas, Denton.

Kline, P. (1998). *The new psychometrics: Science, psychology, and measurement*. London, England: Routledge.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London, England: Routledge.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 105-125). New York, NY: Kluwer.

McCrae, R. R., & Allik, J. (Eds.). (2002). *The five-factor model of personality across cultures*. New York, NY: Kluwer.

McCrae, R. R., & Costa, P. T., Jr. (1989). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. *Multivariate Behavioral Research, 24,* 107-124. doi:10.1207/s15327906mbr2401_7

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509-516. doi:10.1037/0003-066X.52.5.509

McCrae, R. R., & Costa, P. T., Jr. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences, 36,* 587-596. doi:10.1016/S0191-8869(03)00118-1

McCrae, R. R., & Costa, P. T., Jr. (2007). Brief versions of the NEO-PI-3. *Journal of Individual Differences, 28,* 116-128. doi:10.1027/1614-0001.28.3.116

McCrae, R. R., Terracciano, A., & Khoury, B. (2007). Dolce far niente: The positive psychology of personality stability and invariance. In A. Ong & M. Van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 176-188). New York, NY: Oxford University Press.

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

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Miller, E. (1999). Out-of-school youth: A review of Jamaican studies. In A. G. Richardson (Ed.), *Caribbean adolescents and youth: Contemporary issues in personality development and behavior* (pp. 189-229). Brooklyn, NY: Caribbean Diaspora Press.

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Frale, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 407-423). New York, NY: Guilford Press.

Murray, G., Rawlings, D., Allen, N. B., & Trinder, J. (2003). NEO Five-Factor Inventory scores: Psychometric properties in a community sample. *Measurement and Evaluation in Counseling and Development, 36,* 140-150.

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five factor model. *Journal of Research in Personality, 41,* 1213-1233. doi:10.1016/j.jrp.2007.02.003

Muthén, B. O. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association, 74,* 807-811. doi:10.2307/2286403

Muthén, B. O., Kao, C., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement, 28,* 1-22. doi:10.1111/j.1745-3984.1991.tb00340.x

Muthén, B. O., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171-189.

Muthén, L. K., & Muthén, B. O. (2007). Mplus (version 4.2) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P., . . . Wagner, H. (2009). Vegan: Community ecology package [Computer software]. Retrieved from http://vegan.r-forge.r-project.org

Panter, A. T., Swygert, K. A., Dahlstrom, W. G., & Tanaka, J. S. (1997). Factor analytic approaches to personality item-level data. *Journal of Personality Assessment, 68,* 561-589. doi:10.1207/s15327752jpa6803_6

Parker, W. D., & Stumpf, H. (1998). A validation of the five-factor model of personality in academically talented youth across observers and instruments. *Personality and Individual Differences, 25,* 1005-1025. doi:10.1016/S0191-8869(98)00016-6

Planning Institute of Jamaica. (2006). *The transition of Jamaican youth to the world of work*. Kingston, Jamaica: Author.

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

R Development Core Team. (2006). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reise, S. P., Smith, L., & Furr, R. M. (2001). Invariance on the NEO PI-R neuroticism scale. *Multivariate Behavioral Research, 36,* 83-110. doi:10.1207/S15327906MBR3601_04

Renaissance Learning. (2010). *STAR reading computer-adaptive reading test: Technical manual*. Wisconsin Rapids, WI: Author.

Richardson, A. G. (1999a). Adolescent self-esteem: Differences across cultures. In A. G. Richardson (Ed.), *Caribbean adolescents and youth: Contemporary issues in personality development and behaviour* (pp. 5-8). New York, NY: Caribbean Diaspora Press.

Richardson, A. G. (1999b). Personality priorities of Caribbean adolescents. In A. G. Richardson (Ed.), *Caribbean adolescents and youth: Contemporary issues in personality development and behaviour* (pp. 1-4). New York, NY: Caribbean Diaspora Press.

Roberts B. W., & DelVecchio, W. F. (2000). The rank order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126,* 3-25. doi:10.1037/0033-2909.126.1.3

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

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69,* 617-640. doi:10.1111/1467-6494.694157

Rolland, J. P. (2002). The cross-cultural generalizability of the five-factor model of personality. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 7-28). New York, NY: Kluwer Academic/Plenum.

Rolland, J. P., Parker, W. D., & Stumpf, H. (1998). A psychometric examination of the French translation of the NEO-PI-R and NEO-FFI. *Journal of Personality Assessment, 7,* 269-291. doi:10.1207/s15327752jpa7102_13

Roth, M., & Herzberg, P. Y. (2004). A validation and psychometric examination of the Arnett Inventory of Sensation Seeking (AISS) in German adolescents. *European Journal of Psychological Assessment, 20,* 205-215. doi:10.1027/1015-5759.20.3.205

Rushton, J. P., & Irwing, P. (2008). A general factor of personality (GFP) from two meta-analyses of the Big Five: Digman (1997) and Mount, Barrick, Scullen, and Rounds (2005). *Personality and Individual Differences, 45,* 679-683. doi:10.1016/j.paid.2008.07.015

Scandell, D. J. (2000). Development and initial validation of validity scales for the NEO-Five Factor Inventory. *Personality and Individual Differences, 29,* 1153-1162. doi:10.1016/S0191-8869(99)00262-7

Sherry, A., Henson, R. K., & Lewis, J. G. (2003). Evaluating the appropriateness of college-age norms for use with adolescents on the NEO Personality Inventory-Revised. *Assessment, 10,* 71-78. doi:10.1177/1073191102250184

Waller, N. G., Tellegen, A., McDonald, R. P., & Lykken, D. T. (1996). Exploring nonlinear models in personality assessment: Development and preliminary validation of a negative emotionality scale. *Journal of Personality, 64,* 545-576. doi:10.1111/j.1467-6494.1996.tb00521.x

Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development, 34,* 166-176.

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177-203). Mahwah, NJ: Erlbaum.

Yoshimura, K., Ono, Y., Nakamura, K., Nathan, J. H., & Suzuki, K. (2001). Validation of the Japanese version of the NEO Five-Factor Inventory in a large community sample. *Psychological Reports, 88,* 443-449. doi:10.2466/PR0.88.2.443-449