



# Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120



John A. Johnson\*

*Emeritus of Psychology, Pennsylvania State University, United States*

## ARTICLE INFO

### Article history:

Available online 20 May 2014

### Keywords:

Construct validation  
Five-Factor Model  
International Personality Item Pool  
IPIP  
Personality assessment  
Scale development  
Self-reports

## ABSTRACT

The IPIP-NEO (Goldberg, 1999) is a 300-item inventory that measures constructs similar to those in the NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992). Despite evidence for its reliability and validity, the IPIP-NEO is even longer than the original 240-item NEO PI-R. This article details the development of a 120-item version of the IPIP-NEO from an Internet sample ( $N = 21,588$ ) and the subsequent testing of its psychometric properties in Goldberg's (2008) Eugene-Springfield community sample ( $N = 481$ ), two additional large Internet samples ( $Ns = 307,313$  and  $619,150$ ) and a local sample ( $N = 160$ ). Results indicate that the psychometric properties of the 120-item IPIP-NEO compare favorably to the properties of the longer form.

© 2014 Elsevier Inc. All rights reserved.

## 1. Introduction

At the 1996 meeting of the European Conference on Personality, Lewis Goldberg (1999) unveiled a new public-domain resource, the International Personality Item Pool, or IPIP. The IPIP has grown from an initial set of 1252 items to 2413 items, all freely accessible from the IPIP website, <http://ipip.ori.org>. The IPIP website contains not only public domain personality items but also over 300 scales constructed from IPIP items. Some of these scales were designed to serve as proxies for the constructs in commercial inventories, thereby providing a public-domain alternative to these inventories (Goldberg et al., 2006).

One of the first personality measures to be created from the IPIP was a 300-item inventory (Goldberg, 1999) designed to measure constructs similar to those assessed by the 30 facet scales in the NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992). Johnson (2000, 2001) created a version of Goldberg's new inventory that could be administered on the World Wide Web and began referring to Goldberg's 300-item inventory as the IPIP-NEO. Like the NEO PI-R, the IPIP-NEO can yield scores for both the five broad domains of the Five Factor Model (Neuroticism, Extraversion, Conscientiousness, Agreeableness, and Openness to Experience) and also six narrower facets of each broad domain (see Costa & McCrae, 1992).

The first published studies with the IPIP-NEO (Goldberg, 1999) indicated that the scales of this inventory showed a mean alpha reliability of .80, surpassing the mean alpha of .75 for the original

NEO PI-R scales. The IPIP-NEO scales correlated on average  $r = .73$  ( $r = .94$  when corrected for attenuation due to scale unreliability) with the NEO PI-R scales on which they were based. In a comparative validity study, Goldberg (1999) found that the IPIP-NEO predicted health-related behaviors better than the NEO PI-R.

The IPIP-NEO has been translated into Croatian, Danish, Estonian, Finnish, German, Hebrew, Hungarian, Indonesian, Japanese, Romanian, and Slovene (Goldberg, n.d.-b). Studies with the Estonian version (Möttus, Pullmann, & Allik, 2006) affirmed the reliability of the IPIP-NEO scales and their convergent validities with their corresponding NEO PI-R scales. This study also indicated that the readability of IPIP-NEO items is higher than the readability of NEO PI-R items.

Other published research employing the IPIP-NEO has shown it to be useful in studies of topics as diverse as acculturation (Leininger, 2002), anxiety and depression (Lewis et al., 2010; Sutton et al., 2011), cell-phone use (Siddiqui, 2011), competitiveness (Fletcher & Nusbaum, 2008), cortisol levels (Adam et al., 2010; Hauner et al., 2008), health behaviors (Hagger-Johnson & Whiteman, 2007), helping in the workplace (Conway, Rogelberg, & Pitts, 2009), job performance (Griffin & Hesketh, 2004; Wallace & Chen, 2006), leadership style (Ali, Nisar, & Raza, 2011), the startle reflex (Craske et al., 2009), and team performance (Ogot & Okudan, 2006).

Despite the growing evidence for the reliability, validity, and utility of the 300-item IPIP-NEO, this inventory suffers from one major shortcoming: It is even longer than the original 240-item NEO PI-R. Researchers who wish to include an inventory of the five major personality factors within a battery of other psychological

\* Fax: +1 814 375 4784.

E-mail address: [jsj@psu.edu](mailto:jsj@psu.edu)

measures may find the length of the IPIP-NEO prohibitive. Although there are five-factor IPIP inventories containing 20, 50, and 100 items, none of these inventories can measure the six facets within each of the five broad domains. In response to this problem, the author developed a version of the IPIP-NEO that can reliably and validly represent both the five domains and 30 facets of the Five Factor Model with 120 items (four items per facet scale). Study 1 describes the development of this instrument, the IPIP-NEO-120, and Study 2 describes further validation of the IPIP-NEO-120.

## 2. Study 1: Development of four-item facet scales

### 2.1. Method

#### 2.1.1. Participants

The sample was the same Internet sample studied in Johnson (2005),  $N = 21,588$  individuals (7859 male, 13,729 female) who completed, anonymously, the 300-item IPIP-NEO on the World Wide Web between September 3, 1999 and March 18, 2000. A description of the way in which the IPIP-NEO was formatted for administering, scoring, and providing feedback on the Web can be found in Johnson (2000). Participants were not actively recruited; they discovered the Web site on their own or by word-of-mouth. The average age of the sample was 26.2 ( $SD = 10.8$ ). The protocols in the sample met criteria for protocol validity; see Johnson (2005) for details. Prior to screening for protocol validity, 23,994 sets of responses were analyzed. The removal of duplicate protocols and protocols with apparent inattentive responding, too many missing responses, or insufficient internal consistency led to the final sample size of  $N = 21,588$  individuals.

#### 2.1.2. Procedure, results, and discussion

To explain how the IPIP-NEO-120 was constructed, it is helpful to review the three major strategies of scale construction (External, Internal, and Intuitive), as described in a classic monograph by Goldberg (1972). An External strategy selects items based on consistent empirical correlations between responses to the items and other information about the respondents beyond their responses to the items. For example, an External approach to constructing an Extraversion/Introversion scale might begin by having a group of people who know each other well nominate the most extraverted and most introverted individuals in the group. Items that are answered differently by the two groups would constitute a preliminary Extraversion/Introversion scale. Items from this preliminary scale that are again answered differently by a second set of extraverts and introverts (this is called *cross-validation*) would constitute the final scale.

The Internal strategy seeks to identify sets of items that are answered similarly by a group of research participants. Factor analysis has been used to accomplish the aims of the Internal strategy in two ways. First, factor analysis can organize a very large set of diverse items into a limited number of item subsets based on similar endorsement patterns; each of these subsets represents a potential scale. Second, item responses from a potential scale can themselves be factor-analyzed. The number of interpretable factors from such an analysis can indicate whether the scale is measuring one relatively narrow construct or a broad-bandwidth construct that encompasses discernible components (Briggs & Cheek, 1986.) Extraversion, for example, is a broad-bandwidth construct containing subcomponents such as Warmth, Gregariousness, and Assertiveness (Costa & McCrae, 1992). In addition to factor analysis, item analysis (Anastasi, 1976) represents a version of the Internal strategy. Item analysis eliminates items with low correlations with the total score from an item set, which maximizes the internal consistency of the scale.

Intuitive strategies rely on the scale author's professional judgment about the suitability of items for assessing a personality construct. Intuitive strategies are often classified as Rational-Intuitive or Theoretical-Intuitive. The Rational-Intuitive strategy uses common sense about relevance of item content for ordinary-language personality constructs (Wolfe, 1993). For example, "I am a sociable person" would probably be judged as suitable for an Extraversion scale. The straightforwardness of the Rational-Intuitive strategy implies that one does not necessarily have to be a professional psychologist to author a valid personality scale, and evidence supports that suggestion (Ashton & Goldberg, 1973). The Theoretical-Intuitive strategy, in contrast, holds that professional psychologists possess a rarified level of insight about non-commonsense psychological constructs and the kinds of item responses that would assess these theoretical constructs. For example, a Freudian might regard items such as "I am fascinated by fire" and "I like the Washington monument better than the Jefferson memorial" as good items for assessing phallic character (Meehl, 1970).

Although the three major strategies for constructing personality scales can be described as independent techniques, in practice scale authors often combine the strategies. In fact, some form of the Intuitive strategy is usually used to better understand themes within the content of scales created with the External strategy, a process that Gough (1987) called *conceptual analysis*. Furthermore, an Intuitive strategy is absolutely required for interpreting the meaning of scales created with an Internal strategy. Without an intuitive interpretation of item content, possible psychological meanings for a set of item responses that intercorrelate or load highly on a factor remain completely unknowable.

In the development of the 300-item IPIP-NEO, all three strategies were used (Goldberg, n.d.-a). The construction of the IPIP-NEO began by correlating responses to all available items in the International Personality Item Pool (1252 at the time) with the 30 facet scores of the NEO PI-R (Costa & McCrae, 1992). Thus, an External strategy was used as the first step. Items that correlated higher with a particular facet score than any other facet score became candidates for an IPIP version of that facet. The five highest positively correlating and five highest negatively correlating items for each facet became the initial, preliminary scale. However, the aim of identifying equal numbers of positively and negatively keyed items was relaxed if correlations with the original facet scales were substantially higher for items correlating positively or negatively.

Next, a Rational-Intuitive strategy was employed to examine the content of the items. For any pair of items judged to have essentially identical content, only the higher correlating item was retained, and the next highest correlating item became a replacement for the lower correlating item. At this point, the set of items for each facet was examined to see if the content of the item set told "a coherent story" (Goldberg, n.d.-a). Items that did not "mirror the major story-line" (Goldberg, n.d.-a) were omitted and replaced with a new item from the most highly correlating items.

Finally, an Internal strategy was used to see if inclusion of any item lowered the coefficient alpha reliability of the scale. If so, the item was replaced with a new item from the most highly correlating items. This process was repeated to achieve an alpha that was as high as possible without sacrificing breadth of content. Thus, the Internal and Rational-Intuitive strategies were combined at this point in a manner that required "some ingenuity, and thus this is the stage where an exact algorithm would be difficult to formulate" (Goldberg, n.d.-a).

When considering possible methods for constructing four-item IPIP scales for the 30 NEO PI-R facets, I decided to build upon the substantial amount of work that had already been invested into the construction of the 10-item scales. Rather than beginning with an External strategy that identified from the full pool of over 1000 items the four highest correlating items with each facet, I assumed

that the original scale development had already identified 10 excellent IPIP items for each facet. The established reliability and validity of these 10-item scales (Goldberg, 1999) indicate that total scores on these scales are good representations of the 30 constructs and that the quality of individual items can be evaluated with an Internal strategy, correlating individual item responses with the total score. Basing scale construction on the existing 10-item scales also had the advantage of being able to use item responses from over 20,000 individuals who had already completed the 300-item IPIP-NEO online.

Identification of four items for each of the 30 facets proceeded in three phases. The first phase employed the Reliability application from SPSS Base 10.0 (SPSS, 1999) to generate corrected item-total correlations for each of the 10 items in each facet scale. The item with the lowest item-total correlation was identified, removed, and the Reliability application run again until four items remained. Prior to the removal of six items from each facet scale, coefficient alphas for the 10-item facet scales ranged from a low of .72 for E4 Activity level to a high of .92 for N2 Anger. When reduced to 4-item scales, alphas of at least .70 were maintained for every scale except C1 Self-Efficacy (.63), C3 Dutifulness (.69), and O3 Emotionality (.69).

The second phase involved a Rational-Intuitive strategy. I examined the item content of every 4-item scale for three properties: repetitiveness from near-duplicate items, fidelity to the content of items on Costa and McCrae's (1992) original NEO PI-R, and references to disabilities or other areas that might result in legal problems if the inventory were used for personnel selection (U.S. Department of Labor Employment, 1999). I replaced any item whose wording I judged to be too close to one of the other three items with a substitute that maintained the highest level of alpha reliability for two reasons. First, replacing repetitive items broadened the bandwidth of the scales. Second, respondents sometimes object to scales that seem too repetitive (Wolfe, 1993). As an example of a replacement, two of the four items with the highest item-total correlations for O4 Adventurousness, were "Dislike changes" and "Don't like the idea of change." The latter was replaced with "Prefer variety to routine."

As I examined all scales for near-duplicate items, I also attempted to insure that the content of all items was as similar as possible to Costa and McCrae's items. Content similarity, coupled with empirical correlations with the original NEO PI-R scales, increases the probability that evidence for the NEO PI-R's validity would apply to the shortened IPIP-NEO. The only scale with questionable content similarity was O6 Liberalism, which, when compared to its counterpart in the NEO PI-R, O6 Openness to Values, contained a disproportionate number of items dealing with crime, law, and order. It also contained an item referring to belief in one true religion, which is legally problematic (this was the only item with potential legal ramifications). Replacing two items reduced alpha reliability of the four-item O6 facet scale to .64. This lower alpha reliability was considered to be an acceptable trade-off for avoiding repetitiveness and potential legal problems.

The third phase of scale development included the computation of alphas for the five domain scales and the computation of alphas for domain and facet scales separately by sex. Alphas for each sex were found to be nearly identical to alphas for the full sample for most scales. Table 1 presents the final items for each shortened scale with the original item numbers from the 300-item inventory and the alphas and item-total correlations for the full sample. Although a few scales did not reach the common, rule-of-thumb standard of alpha reliability acceptability, .70, all alphas were at least .60, a level often considered adequate for four-item scales in the development stage of research (Hair, Black, Babin, Anderson, & Tatham, 2006).

Questions could be raised about whether the scale construction strategy used here was optimal. Goldberg (1972) found little

difference in the validities of scales constructed by the External, Internal, and Intuitive strategies from the same pool of items. The major influences on validity in the Goldberg (1972) study were found to be the predictability of the criterion behaviors (e.g., rated sociability was much more predictable than choice of major) and a criterion predictability  $\times$  construction strategy interaction (Internal and Intuitive scales showed cross-validities averaging around  $r = .40$  for the most predictable criteria but only around  $r = .10$  for the least predictable criteria, while External scales correlated about .30 with the most predictable criteria and around .20 with the least predictable criteria). If the findings of Goldberg (1972) generalize to the IPIP, there is no reason to believe that a purely External, Internal, or Intuitive strategy would have produced markedly more valid scales across the full range of predictable and less predictable criteria. Researchers who think otherwise are free to conduct a study with the IPIP similar to the Goldberg (1972) study.

Since the publication of Goldberg (1972), a new personality scale construction tool, item response theory (IRT), has been forwarded as a way of improving upon traditional scale construction strategies (Morizot, Ainsworth, & Reise, 2007). Morizot et al. (2007) note that IRT is not a replacement for the traditional methods of scale construction based on classical test theory (CTT), "It is important to conduct these basic analyses [inter-item and item-total correlations; internal consistency] because if some items show poor psychometric properties with CTT statistics, they will almost certainly not be good for IRT modeling either" (p. 412). Similarly, Reeve and Fayers (2005) write, "Applying IRT models does not imply abandoning CTT. Rather, IRT complements CTT to provide thorough analysis of an instrument." (p. 69). Among the things that IRT adds to traditional methods is the ability to see how well items discriminate across different levels of a trait. Sibley (2012), for example, demonstrated with IRT that four-item IPIP scales representing the five major personality domains plus Honesty-Humility (Sibley et al., 2011) were "reasonably precise short-form measures of each of the six major broad-bandwidth dimensions of personality across a fairly broad range of each latent trait centered on average or mean levels of each trait" (p. 26).

In the present study, IRT could not have been used by itself to construct 30 four-item facet scales straight from the full IPIP or even the existing 300-item IPIP-NEO, because IRT assumes unidimensionality for a set of items under analysis. That is, it is assumed that one underlying factor accounts for a person's response to an item (Morizot et al., 2007; Reeve & Fayers, 2005). In fact, Morizot et al. (2007) recommend conducting an exploratory factor analysis upon a set of items prior to employing IRT to assess the unidimensionality of that set of items. Obviously, the full IPIP and 300-item IPIP-NEO are multidimensional. Therefore, at best, IRT could have been used in the present study to select four items from each existing 10-item facet scale. I decided to forgo labor-intensive IRT techniques for the simpler Internal/Intuitive strategy on the hunch that neither technique was more likely to identify substantially more valid four-item scales from each 10-item pool. Future research could test this hunch.

One final scale construction issue that needs further explanation is the decision not to select two positively-keyed and two negatively-keyed items for each scale. Some treatises on personality scale construction (e.g., Furr, 2011; Wolfe, 1993) recommend that scales be "balanced" by including a roughly equal number of positively-keyed and negatively keyed items. The assumption underlying balancing positively-keyed and negatively-keyed items is that this reduces the effects of hypothetical response biases toward "yea-saying" (often called *acquiescence response bias*) and "nay-saying" (Furr, 2011). I decided not to include item balance as a standard for item inclusion for two reasons. First, maximizing alpha reliability was considered to be more important than item balance, and achieving alpha levels of at least .70 sometimes

**Table 1**

Item assignments and alphas for IPIP-NEO-120 Scales.

IPIP-300 item no.	IPIP-120 item no.	Facet key	Domain scales, facet scales, and items	Alphas <sup>a</sup> & item-total <i>r</i> s
<i>Neuroticism</i>				<b>.90</b>
			N1 Anxiety	<b>.78</b>
1	1	+N1	Worry about things	.58
31	31	+N1	Fear for the worst	.60
61	61	+N1	Am afraid of many things	.57
91	91	+N1	Get stressed out easily	.61
			N2 Anger	<b>.87</b>
6	6	+N2	Get angry easily	.76
36	36	+N2	Get irritated easily	.72
126	66	+N2	Lose my temper	.75
216	96	–N2	Am not easily annoyed	.65
			N3 Depression	<b>.85</b>
11	11	+N3	Often feel blue	.69
41	41	+N3	Dislike myself	.70
71	71	+N3	Am often down in the dumps	.76
251	101	–N3	Feel comfortable with myself	.63
			N4 Self-Consciousness	<b>.74</b>
76	16	+N4	Find it difficult to approach others	.61
106	46	+N4	Am afraid to draw attention to myself	.52
136	76	+N4	Only feel comfortable with friends	.52
256	106	–N4	Am not bothered by difficult social situations	.46
			N5 Immoderation	<b>.72</b>
111	21	+N5	Go on binges	.47
171	51	–N5	Rarely overindulge	.54
201	81	–N5	Easily resist temptations	.49
231	111	–N5	Am able to control my cravings	.55
			N6 Vulnerability	<b>.76</b>
26	26	+N6	Panic easily	.61
56	56	+N6	Become overwhelmed by events	.57
86	86	+N6	Feel that I'm unable to deal with things	.54
176	116	–N6	Remain calm under pressure	.53
<i>Extraversion</i>				<b>.89</b>
			E1 Friendliness	<b>.81</b>
2	2	+E1	Make friends easily	.61
62	32	+E1	Feel comfortable around people	.68
212	62	–E1	Avoid contacts with others	.66
272	92	–E1	Keep others at a distance	.58
			E2 Gregariousness	<b>.79</b>
7	7	+E2	Love large parties	.67
37	37	+E2	Talk to a lot of different people at parties	.61
157	67	–E2	Prefer to be alone	.48
247	97	–E2	Avoid crowds	.64
			E3 Assertiveness	<b>.85</b>
12	12	+E3	Take charge	.76
42	42	+E3	Try to lead others	.70
132	72	+E3	Take control of things	.72
162	102	–E3	Wait for others to lead the way	.61
			E4 Activity level	<b>.71</b>
17	17	+E4	Am always busy	.63
47	47	+E4	Am always on the go	.60
77	77	+E4	Do a lot in my spare time	.52
167	107	–E4	Like to take it easy	.26
			E5 Excitement Seeking	<b>.77</b>
22	22	+E5	Love excitement	.58
52	52	+E5	Seek adventure	.59
142	82	+E5	Enjoy being reckless	.57
172	112	+E5	Act wild and crazy	.60
			E6 Cheerfulness	<b>.80</b>
27	27	+E6	Radiate joy	.58
57	57	+E6	Have a lot of fun	.59
147	87	+E6	Love life	.64
177	117	+E6	Look at the bright side of life	.66
<i>Openness to experience</i>				<b>.83</b>
			O1 Imagination	<b>.76</b>
3	3	+O1	Have a vivid imagination	.45
33	33	+O1	Enjoy wild flights of fantasy	.56
63	63	+O1	Love to daydream	.66
93	93	+O1	Like to get lost in thought	.56

(continued on next page)

Table 1 (continued)

IPIP-300 item no.	IPIP-120 item no.	Facet key	Domain scales, facet scales, and items	Alphas <sup>a</sup> & item-total rs
			O2 Artistic interests	<b>.76</b>
8	8	+O2	Believe in the importance of art	.63
68	38	+O2	See beauty in things that others might not notice	.44
188	68	−O2	Do not like poetry	.54
218	98	−O2	Do not enjoy going to art museums	.62
			O3 Emotionality	<b>.69</b>
13	13	+O3	Experience my emotions intensely	.49
43	43	+O3	Feel others' emotions	.44
223	73	−O3	Rarely notice my emotional reactions	.46
283	103	−O3	Don't understand people who get emotional	.52
			O4 Adventurousness	<b>.72</b>
18	18	+O4	Prefer variety to routine	.43
138	48	−O4	Prefer to stick with things that I know	.53
168	78	−O4	Dislike changes	.58
288	108	−O4	Am attached to conventional ways	.49
			O5 Intellect	<b>.75</b>
53	23	+O5	Love to read challenging material	.43
203	53	−O5	Avoid philosophical discussions	.61
233	83	−O5	Have difficulty understanding abstract ideas	.52
263	113	−O5	Am not interested in theoretical discussions	.65
			O6 Liberalism	<b>.64</b>
28	28	+O6	Tend to vote for liberal political candidates	.54
58	58	+O6	Believe that there is no absolute right or wrong	.31
148	88	−O6	Tend to vote for conservative political candidates	.55
268	118	−O6	Believe that we should be tough on crime	.31
<i>Agreeableness</i>				<b>.87</b>
			A1 Trust	<b>.86</b>
4	4	+A1	Trust others	.72
34	34	+A1	Believe that others have good intentions	.65
64	64	+A1	Trust what people say	.76
184	94	−A1	Distrust people	.74
			A2 Morality	<b>.76</b>
99	9	−A2	Use others for my own ends	.59
159	39	−A2	Cheat to get ahead	.54
249	69	−A2	Take advantage of others	.68
279	99	−A2	Obstruct others' plans	.47
			A3 Altruism	<b>.76</b>
74	14	+A3	Love to help others	.58
104	44	+A3	Am concerned about others	.65
194	74	−A3	Am indifferent to the feelings of others	.54
284	104	−A3	Take no time for others	.52
			A4 Cooperation	<b>.73</b>
169	19	−A4	Love a good fight	.46
199	49	−A4	Yell at people	.53
229	79	−A4	Insult people	.58
259	109	−A4	Get back at others	.53
			A5 Modesty	<b>.76</b>
144	24	−A5	Believe that I am better than others	.52
174	54	−A5	Think highly of myself	.71
204	84	−A5	Have a high opinion of myself	.71
264	114	−A5	Boast about my virtues	.32
			A6 Sympathy	<b>.72</b>
29	29	+A6	Sympathize with the homeless	.58
59	59	+A6	Feel sympathy for those who are worse off than myself	.58
149	89	−A6	Am not interested in other people's problems	.41
239	119	−A6	Try not to think about the needy	.49
<i>Conscientiousness</i>				<b>.90</b>
			C1 Self-Efficacy	<b>.63</b>
5	5	+C1	Complete tasks successfully	.50
35	35	+C1	Excel in what I do	.46
65	65	+C1	Handle tasks smoothly	.50
155	95	+C1	Know how to get things done	.21
			C2 Orderliness	<b>.83</b>
40	10	+C2	Like to tidy up	.61
160	40	−C2	Often forget to put things back in their proper place	.63
190	70	−C2	Leave a mess in my room	.72
220	100	−C2	Leave my belongings around	.67



Table 1 (continued)

IPIP-300 item no.	IPIP-120 item no.	Facet key	Domain scales, facet scales, and items	Alphas <sup>a</sup> & item-total rs
			C3 Dutifulness	<b>.69</b>
45	15	+C3	Keep my promises	.53
105	45	+C3	Tell the truth	.52
165	75	–C3	Break rules	.33
195	105	–C3	Break my promises	.58
			C4 Achievement-striving	<b>.80</b>
50	20	+C4	Work hard	.63
140	50	+C4	Do more than what's expected of me	.57
260	80	–C4	Do just enough work to get by	.67
290	110	–C4	Put little time and effort into my work	.64
			C5 Self-Discipline	<b>.73</b>
55	25	+C5	Am always prepared	.45
145	55	+C5	Carry out my plans	.52
205	85	–C5	Waste my time	.57
265	115	–C5	Have difficulty starting tasks	.58
			C6 Cautiousness	<b>.87</b>
120	30	–C6	Jump into things without thinking	.76
150	60	–C6	Make rash decisions	.69
210	90	–C6	Rush into things	.72
270	120	–C6	Act without thinking	.75

<sup>a</sup> Cronbach alpha coefficients in boldface.

required using all positively- or negative-keyed items. Second, I've seen little convincing evidence that response biases play an important role in personality measurement. Rorer (1965) argued effectively against response biases long ago. Furthermore, Schriesheim and Hill (1981) note, "The prevailing conventional wisdom is that it is advisable to mix positively and negatively worded items in psychological measures to counteract acquiescence response bias. However, there has been virtually no unambiguous empirical evidence to support this recommendation" (p. 1101). Schriesheim and Hill (1981) demonstrated that employing negatively-keyed items to reduce acquiescence can actually impair response accuracy.

As a consequence of not using balanced keying as a criterion for item selection, the proportions of positively- and negatively-keyed items in the IPIP-NEO-120 are not as equal as in the 300-item IPIP-NEO or in the NEO PI-R. Proportions of positively-keyed items in the IPIP-NEO-120 domain scales are as follows: Neuroticism, 71%; Extraversion, 75%; Openness to Experience, 50%; Agreeableness, 29%; Conscientiousness, 46%. Overall, there are 65 positively-keyed and 55 negatively-keyed items. In comparison, positively-keyed items in the 300-item IPIP-NEO represent 55%, 60%, 47%, 40%, and 53% of items for the N, E, O, A, and C domain scales, respectively, or 148 out of 300 items. Positively-keyed items in the 240-item NEO PI-R represent 56%, 60%, 50%, 54%, and 58% of N, E, O, A, and C items, respectively, or 134 out of 240 items.

Whether or not keying imbalance in the short IPIP-NEO scales lowers their validities due to acquiescence was tested in Study 2 by creating an Acquiescence Index (AI) and computing ipsatized scores with the procedure described in the Appendix A of Soto, John, Gosling, and Potter (2008) and then examining whether these ipsatized scores produced a clearer factor structure (Rammstedt, Goldberg, & Borg, 2010) or correlated higher with external criteria than normal scores.

### 3. Study 2: Further validation of the IPIP-NEO-120

#### 3.1. Participants and measures

Four samples were used to further assess the psychometric properties of the IPIP-NEO-120. The first was Goldberg's (2008) Eugene-Springfield community sample ( $N = 481$ ), which completed a paper-and-pencil version of the 300-item IPIP-NEO and Costa and McCrae's (1992) NEO PI-R; 420 of these participants were also

described by acquaintances with the Big Five Inventory (BFI; John, Naumann, & Soto, 2008) and the Big Five Mini-Marker scales (Saucier, 1994). The 300-item IPIP-NEO (henceforth called the IPIP-NEO-300) was scored for both the original 10-item facet scales and 4-item facet scales to compare the reliabilities of the longer and shorter forms and the relative abilities of both forms to predict NEO PI-R scores and acquaintance ratings of the five major personality factors.

Two Internet samples were used. Over the past 12 years, more than 300,000 persons have anonymously completed an online version of the IPIP-NEO-300 and more than 600,000 persons have completed the IPIP-NEO-120 at the author's Web site, <http://www.personal.psu.edu/~j5j/IPIP/>. The primary purpose of the Web site is to educate respondents about the Five-Factor Model of personality; respondents receive narrative feedback upon completing either inventory. Johnson's (2005) criteria for protocol validity were used to remove duplicate protocols and protocols with apparent inattentive responding, too many missing responses, or insufficient internal consistency. Prior to screening for protocol validity, sample sizes were 334,161 cases for the IPIP-NEO-300 and 690,863 cases for the IPIP-NEO-120. After screening, final  $N$ s were 307,313 (122,164 male, 185,149 female) for the IPIP-NEO-300 and 619,150 (248,258 male, 370,892 female) for the IPIP-NEO-120. The average age for persons completing the IPIP-NEO-300 was 25.2 ( $SD = 10.0$ ); for the IPIP-NEO-120 the average age was also 25.2 ( $SD = 10.2$ ). Alpha reliabilities and factor structure for the 10-item IPIP-NEO-300 scales, the 4-item scales scored from the IPIP-NEO-300, and the scales of the IPIP-NEO-120 were compared.

The fourth sample contained  $N = 160$  participants (59 males, 101 females) who completed the IPIP-NEO-300 online and were judged by acquaintances on descriptions of the five domains and 30 facets of personality (Johnson, 2009). Correlations between these acquaintance ratings and the long and short scales scored from the IPIP-NEO-300 provide additional primary evidence of the relative validity of the long and short IPIP-NEO scales.

#### 3.2. Results and discussion

##### 3.2.1. Reliability

Alpha reliability coefficients from the Eugene-Springfield community sample and two Internet samples are displayed in Table 2. Data from the community sample and first Internet sample show,

**Table 2**

Alpha reliability coefficients for the IPIP-NEO inventories.

Scale labels		Eugene-spring field sample (N = 501)		Internet sample (N = 307,313)		Internet sample (N = 619,150)
IPIP-NEO	NEO PI-R	IPIP 300	IPIP 120	IPIP 300	IPIP 120	IPIP 120
<i>Neuroticism</i>	<i>N</i>	.94	.88	.95	.90	.90
Anxiety	N1	.83	.71	.86	.78	.78
Anger	N2	.88	.77	.91	.86	.87
Depression	N3	.89	.80	.91	.86	.85
Self-consciousness	N4	.80	.63	.82	.72	.70
Immoderation	N5	.77	.69	.77	.71	.69
Vulnerability	N6	.82	.70	.85	.76	.76
<i>Extraversion</i>	<i>E</i>	.92	.84	.94	.89	.89
Friendliness	E1	.87	.77	.88	.81	.81
Gregariousness	E2	.79	.60	.88	.79	.79
Assertiveness	E3	.84	.75	.85	.83	.85
Activity level	E4	.71	.68	.71	.70	.69
Excitement-seeking	E5	.77	.67	.84	.75	.73
Cheerfulness	E6	.81	.71	.82	.79	.79
<i>Openness to experience</i>	<i>O</i>	.92	.85	.90	.82	.81
Imagination	O1	.82	.70	.84	.75	.74
Artistic interests	O2	.85	.72	.80	.74	.74
Emotionality	O3	.81	.67	.77	.66	.65
Adventurousness	O4	.77	.66	.80	.70	.70
Intellect	O5	.86	.78	.84	.74	.73
Liberalism	O6	.86	.76	.77	.64	.63
<i>Agreeableness</i>	<i>A</i>	.90	.81	.92	.85	.86
Trust	A1	.82	.70	.88	.86	.85
Morality	A2	.74	.62	.78	.74	.74
Altruism	A3	.77	.65	.82	.74	.73
Cooperation	A4	.72	.56	.77	.70	.71
Modesty	A5	.76	.63	.77	.75	.73
Sympathy	A6	.75	.68	.77	.70	.72
<i>Conscientiousness</i>	<i>C</i>	.92	.84	.94	.90	.90
Self-efficacy	C1	.79	.57	.81	.63	.77
Orderliness	C2	.83	.76	.85	.82	.83
Dutifulness	C3	.71	.47	.78	.67	.67
Achievement-striving	C4	.79	.68	.82	.79	.79
Self-discipline	C5	.85	.66	.89	.73	.71
Cautiousness	C6	.76	.70	.84	.86	.88

as one would expect, that the alphas for the 4-item facet scales of the IPIP-NEO-120 are slightly lower than the alphas for the 10-item facet scales (mean alphas of .80 vs. .68 in the community sample and .82 vs. .75 in the Internet samples. Perhaps of most interest are the alpha coefficients in the larger Internet sample ( $N = 619,150$ ) because scores from this sample come directly from the 120-item inventory rather than the 300-item inventory scored for the 120-item version. Facet score alphas range from .63 to .88, with all but three facets showing alphas of .69 or greater. Although .70 is often cited as a minimum level of acceptable personality scale reliability, in reality “sufficient” reliability depends upon the purpose of psychological testing. Nunnally (1978) suggests that .70 is sufficient in the early stages of research, while .95 is a desirable standard when making important life decisions about individuals. Hair et al. (2006) say that alphas above .60 are acceptable for research, especially if the scales have only a few items. By these standards, the IPIP-NEO-120 facet scales have sufficient reliability for research studies, but probably should not be used to make important decisions about individuals.

### 3.2.2. Primary validity

The original IPIP-NEO was designed to measure constructs similar to those in the NEO PI-R (Costa & McCrae). Therefore, the primary validity of the IPIP-NEO inventories is represented by the correlations between its scales and the corresponding scales of the NEO PI-R. Those correlations, shown in Table 3, average .73 (.94 corrected for attenuation due to unreliability) for the longer scales from the IPIP-NEO-300 and .66 (.91 corrected for

attenuation) for the 4-item scales from the IPIP-NEO-120. Although the long and short IPIP-NEO scales do not measure exactly the same constructs as the NEO PI-R, they measure highly similar constructs, and correlations with the NEO PI-R scales are only slightly lower for the IPIP-NEO-120 4-item scales.

### 3.2.3. Correlations between scale scores and acquaintance ratings

Additional evidence for the validity of the long and short IPIP-NEO scales can be found in the correlations between these scales and corresponding acquaintance ratings of the five major personality factors as assessed by the BFI and Mini-Markers. These values, computed from the Eugene-Springfield community sample, are also shown in Table 3. The correlations with these measures were expected to be lower than with the NEO PI-R because of method differences (self-report vs. acquaintance ratings) and the different operationalizations of the five factors in the BFI and Mini-Markers. Also, the facet scales represent narrower subdomains of the five personality factors and therefore should not correlate as highly as the five domain scores.

Table 3 shows that the self-acquaintance correlations are well within expectable values for self/other agreement, given the factors that affect the accuracy of self-perception and the perception of others (Funder, 1999; Vazire, 2010). The table also shows that the magnitude of correlations with the acquaintance ratings was nearly identical for the 10-item and 4-item facet scales. For the five IPIP-NEO domain scales, correlations with the corresponding five BFI scales averaged .48 and .46 and with the five Mini-Marker scales, .43 and .40, for the long and short IPIP-NEO scales,

**Table 3**

Validity of the IPIP-NEO inventories.

Scale labels		Eugene-springfield sample (Ns = 420–501)						Local sample (N = 160)	
		Correlations with NEO		Correlations with BFI <sup>a</sup>		Correlations with MiniMarkers <sup>a</sup>		Correlations with acquaintance ratings <sup>b</sup>	
IPIP-NEO	NEO PI-R	IPIP 300	IPIP 120	IPIP 300	IPIP 120	IPIP 300	IPIP 120	IPIP 300	IPIP 120
Neuroticism	N	.88	.87	.47	.46	.34	.33	.52	.49
Anxiety	N1	.76	.76	.45	.44	.32	.30	.44	.40
Anger	N2	.77	.71	.41	.32	.36	.29	.55	.55
Depression	N3	.81	.76	.39	.38	.30	.28	.61	.61
Self-consciousness	N4	.73	.60	.25	.18	.10	.06	.47	.33
Immoderation	N5	.74	.65	.24	.25	.24	.21	.33	.33
Vulnerability	N6	.78	.74	.36	.36	.22	.23	.44	.43
Extraversion	E	.89	.85	.51	.49	.48	.45	.43	.42
Friendliness	E1	.76	.68	.43	.42	.42	.41	.41	.39
Gregariousness	E2	.78	.73	.36	.37	.33	.35	.42	.37
Assertiveness	E3	.81	.73	.45	.29	.46	.30	.47	.39
Activity level	E4	.72	.63	.34	.31	.31	.27	.37	.36
Excitement-seeking	E5	.67	.59	.15	.15	.12	.10	.46	.43
Cheerfulness	E6	.77	.69	.34	.28	.29	.24	.39	.42
Openness to experience	O	.87	.84	.58	.57	.52	.49	.30	.27
Imagination	O1	.74	.69	.37	.29	.34	.29	.26	.20
Artistic interests	O2	.80	.76	.45	.45	.37	.34	.36	.36
Emotionality	O3	.71	.65	.34	.33	.33	.25	.42	.39
Adventurousness	O4	.72	.62	.40	.38	.28	.30	.28	.19
Intellect	O5	.81	.75	.43	.41	.31	.40	.24	.23
Liberalism	O6	.71	.63	.40	.36	.44	.33	.35	.34
Agreeableness	A	.83	.76	.38	.33	.35	.31	.29	.25
Trust	A1	.78	.73	.29	.31	.26	.27	.30	.28
Morality	A2	.65	.54	.22	.15	.22	.16	.14	.09
Altruism	A3	.68	.54	.32	.23	.28	.19	.16	.14
Cooperation	A4	.72	.62	.40	.31	.35	.29	.25	.25
Modesty	A5	.71	.64	.10	.08	.08	.07	.22	.24
Sympathy	A6	.62	.55	.21	.18	.23	.21	.44	.35
Conscientiousness	C	.84	.80	.46	.45	.45	.44	.42	.42
Self-efficacy	C1	.68	.59	.22	.25	.21	.24	.26	.25
Orderliness	C2	.77	.68	.45	.41	.49	.43	.57	.52
Dutifulness	C3	.60	.53	.32	.27	.27	.24	.35	.33
Achievement-striving	C4	.71	.57	.32	.35	.30	.33	.37	.41
Self-discipline	C5	.77	.72	.31	.31	.30	.31	.37	.32
Cautiousness	C6	.69	.61	.32	.21	.28	.19	.28	.27

<sup>a</sup> Correlations of  $r = .0804$  or greater are significant at the  $p < .05$  level (two-tailed).<sup>b</sup> Correlations of  $r = .1306$  or greater are significant at the  $p < .05$  level (two-tailed).

respectively. The strongest correspondence between self and acquaintance scores was for Openness to Experience ( $r$ s ranging from .49 to .58), and the weakest, for Agreeableness ( $r$ s from .31 to .38). As expected, the narrower facet scales correlated slightly lower with the BFI domain scales, averaging .33 and .30, and with the Mini-Marker domain scales, averaging .29 and .27 for the long and short IPIP-NEO scales, respectively. As with the correlations between IPIP domain scales and corresponding BFI and Mini-Marker scales, the strongest correspondence was found for the Openness to Experience facets (average  $r$ s between .32 and .40), and the weakest, for Agreeableness (average  $r$ s between .20 and .26). Studies of self-observer agreement typically show the lowest levels of correspondence for the Agreeableness domain, possibly because of the unique perceptions and motivations of actors and observers (see Edmonds, Goldberg, Hampson, & Barckley, 2013).

Acquaintance ratings in the local sample ( $N = 160$ ) are of particular interest because acquaintances rated targets who had completed the IPIP-NEO-300 on written descriptions of the five domains and all 30 facets (Johnson, 2009). For example, for the Friendliness facet, acquaintances were asked to rate the standing of the target on friendliness with the following description: "Friendliness. Friendly people genuinely like other people and openly demonstrate positive feelings toward others. They make friends quickly and it is easy for them to form close, intimate relationships. Low levels of Friendliness do not necessarily indicate

coldness and hostility, but people with low Friendliness do not reach out to others and are perceived as distant and reserved." Raters were asked to rate targets on an 11-point percentile scale, with the rating categories labeled as follows: 1 or 10, Low; 20, Low-Average; 30, 40, 50, 60, or 70, Average; 80, High-Average; 90 or 99, Very High. The Johnson (2009) study was designed to examine the correspondence between standardized self-report scores and acquaintance perceptions of low, average, or high trait standing; for the current study we are interested simply in the correlations between IPIP-NEO self-report scores and the acquaintance ratings of the corresponding domains and facets.

For the five broad domains, the long and short IPIP-NEO scales showed about equally good correspondence with acquaintance ratings. The average correlation for the 10-item scales was .39, and for the 4-item scales, .37. Neuroticism showed the strongest agreement ( $r$ s = .52 and .49 for the long and short scales, respectively), and Agreeableness once again showed the weakest Agreement ( $r$ s = .29 and .25, respectively). At the facet level, again the long and short versions of the IPIP-NEO showed comparable correspondence with acquaintance ratings, with average  $r$ s of .37 and .34, respectively. Paralleling findings for the domains, the Neuroticism facets showed the strongest correspondence (average  $r$ s = .47 and .44) and Agreeableness facets, the weakest correspondence (average  $r$ s = .25 and .23) with the acquaintance ratings. Differences in the magnitudes of correlations across personality factors might



be understood in terms of differences in motivation and information available to judges and targets (Funder, 1999; Vazire, 2010); for the purpose of the current study the important finding is that the short IPIP-NEO scales are working about as well as the original, longer scales.

### 3.2.4. Factor structure

The two large Internet samples were used to examine the factor structure of the IPIP-NEO-300 and IPIP-NEO-120. Three principal components analyses were conducted. In each case, five components were extracted and subjected to varimax rotation with Kaiser normalization. The first analysis entered scores from the 30 10-item facet scales, the second used 4-item facet scale scores computed from the IPIP-NEO-300, and the third used 4-item facet scale scores from the IPIP-NEO-120. Results are displayed in Table 4.

Although the usual five major personality factors are clearly represented by the loadings in all three analyses, not all facets show their highest loading on the expected component. For example, in all three analyses, N4 Self-Consciousness shows a stronger (negative) loading on the Extraversion factor than the Neuroticism factor. The difference in loadings was less pronounced for the analysis of the 10-item facet scales (−.54 and .51 for E and N, respectively) than for the 4-item scales scored from the IPIP-NEO-300 (−.71 and .23) and the 4-item scales from the IPIP-NEO-120 (−.73 and .00). A similar pattern can be found for E4 Activity Level, which loads higher on the Conscientiousness

factor than the Extraversion factor, and O3 Emotionality, which loads higher on the Neuroticism factor than the Openness to Experience factor.

If the original 10-item IPIP-NEO facet scales do not show their highest loadings on the expected factor, it is unlikely that the 4-item scales will show their highest loadings on the expected factor. But, in some cases, the IPIP-NEO-120 facets scales fail to show their expected highest loadings even when the 10-item facets scales do. Examples are the 4-item E3 Assertiveness scale, which loads higher on the Conscientiousness factor than the Extraversion factor, and the 4-item A4 Cooperation scale and 4-item C6 Cautiousness scale, which load highest on the Neuroticism factor (opposite in sign from the Neuroticism facet loadings).

An important question is whether the failure of all facets to load primarily on their expected factor is due to imbalance in the IPIP scales coupled with acquiescent responding from participants, or to the IPIP factor structure simply reflecting the structure of its parent inventory, the NEO PI-R. Because some of the IPIP-NEO-120 facet scales do not contain equal numbers of positively- and negatively-keyed items, scores from respondents who are prone to acquiescence response bias could be inappropriately high or low, possibly distorting the factor structure of the inventory (Rammstedt et al., 2010). To test for this possibility, a content-balanced tendency toward acquiescence indicator (Acquiescence Index or AI) was created by computing the mean of responses to 29 pairs of items with opposite implications for personality (e.g., item 4, “Trust others” vs. item 94, “Distrust people.”). Responses

**Table 4**  
Factor component loadings of IPIP-NEO facet scale scores.

Scale labels	IPIP-NEO	Component loadings, IPIP-NEO-300 10-item scales					Component loadings, IPIP-NEO-300 4-item scales					Component loadings, IPIP-NEO-120 4-item scales				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
<i>Neuroticism</i>	<i>N</i>															
Anxiety	N1	−.27	−.03	<b>.86</b>	.04	−.03	−.33	−.09	<b>.76</b>	.15	−.12	−.52	−.54	−.08	.35	−.20
Anger	N2	−.11	.03	<b>.74</b>	−.40	−.03	−.08	.00	<b>.73</b>	−.29	−.04	−.30	−.72	.09	−.06	−.12
Depression	N3	−.38	−.32	<b>.69</b>	−.10	.08	−.49	−.29	<b>.59</b>	.04	.08	−.63	−.40	−.28	.17	.05
Self-consciousness	N4	−.54	−.30	<b>.51</b>	.22	−.12	−.71	−.21	.23	.12	−.10	−.73	.00	−.24	.09	−.12
Immoderation	N5	.21	−.34	<b>.53</b>	−.21	.08	.10	−.39	<b>.44</b>	−.07	.13	−.01	−.52	−.28	.07	.09
Vulnerability	N6	−.16	−.30	<b>.82</b>	.06	−.12	−.29	−.25	<b>.70</b>	.18	−.16	−.47	−.52	−.23	.34	−.23
<i>Extraversion</i>	<i>E</i>															
Friendliness	E1	<b>.82</b>	.20	−.15	.24	−.03	<b>.81</b>	.17	−.14	.22	−.07	<b>.82</b>	.02	.19	.20	−.07
Gregariousness	E2	<b>.85</b>	.07	−.06	.02	−.07	<b>.83</b>	.02	.02	.05	−.04	<b>.79</b>	−.21	.08	.10	−.06
Assertiveness	E3	<b>.59</b>	<b>.48</b>	−.13	−.34	.19	<b>.45</b>	<b>.56</b>	.00	−.26	.17	.37	−.15	<b>.62</b>	−.16	.17
Activity level	E4	.28	<b>.67</b>	.01	−.15	.02	.32	<b>.60</b>	.10	.02	−.02	.22	−.07	<b>.62</b>	.11	−.04
Excitement-seeking	E5	<b>.65</b>	−.18	.00	−.35	.17	<b>.63</b>	−.23	.22	−.22	.23	<b>.53</b>	−.49	−.13	−.11	.22
Cheerfulness	E6	<b>.73</b>	.04	−.17	.19	.14	<b>.70</b>	.18	−.26	.22	.01	<b>.75</b>	.13	.19	.17	−.01
<i>Openness to experience</i>	<i>O</i>															
Imagination	O1	.07	−.18	.15	−.06	<b>.68</b>	.07	−.25	.21	.00	<b>.58</b>	.00	−.33	−.21	.11	<b>.54</b>
Artistic interests	O2	.20	.11	.11	.30	<b>.63</b>	−.01	.10	.04	.27	<b>.71</b>	−.03	.04	.09	.34	<b>.69</b>
Emotionality	O3	.21	.18	<b>.56</b>	.24	<b>.48</b>	.14	.13	<b>.42</b>	<b>.50</b>	.31	.00	−.20	.13	<b>.68</b>	.23
Adventurousness	O4	<b>.40</b>	.05	−.31	−.04	<b>.53</b>	.38	−.01	−.24	−.04	<b>.54</b>	<b>.40</b>	.09	.01	−.05	<b>.56</b>
Intellect	O5	−.10	.21	−.23	−.07	<b>.79</b>	−.10	.15	−.16	−.02	<b>.78</b>	−.06	.14	.17	.01	<b>.78</b>
Liberalism	O6	−.08	−.31	−.02	.02	<b>.55</b>	−.02	−.23	−.03	.06	<b>.47</b>	−.02	−.05	−.22	.05	<b>.45</b>
<i>Agreeableness</i>	<i>A</i>															
Trust	A1	<b>.45</b>	.01	−.26	<b>.53</b>	.02	<b>.40</b>	−.05	−.27	<b>.48</b>	−.07	<b>.47</b>	.29	−.08	.35	−.07
Morality	A2	−.13	.25	−.06	<b>.77</b>	−.02	−.13	.29	−.22	<b>.65</b>	.02	−.02	<b>.54</b>	.19	<b>.50</b>	.00
Altruism	A3	<b>.45</b>	.22	.03	<b>.70</b>	.18	.28	.18	.10	<b>.73</b>	.18	.24	.13	.17	<b>.77</b>	.11
Cooperation	A4	−.05	−.04	−.22	<b>.81</b>	−.04	−.15	.13	−.43	<b>.64</b>	.00	.01	<b>.70</b>	.02	<b>.41</b>	.02
Modesty	A5	−.34	−.20	.20	<b>.49</b>	−.21	−.27	−.20	.17	<b>.48</b>	−.14	−.25	.16	−.28	<b>.46</b>	−.17
Sympathy	A6	.15	−.03	.18	<b>.70</b>	.34	.15	.07	.13	<b>.67</b>	.26	.11	.08	.04	<b>.71</b>	.20
<i>Conscientiousness</i>	<i>C</i>															
Self-efficacy	C1	.16	<b>.68</b>	−.47	.07	.23	.13	<b>.72</b>	−.23	.11	.11	.21	.20	<b>.74</b>	−.03	.08
Orderliness	C2	−.13	<b>.66</b>	.08	.18	−.29	−.05	<b>.58</b>	−.08	.10	−.23	−.04	.25	<b>.52</b>	.07	−.22
Dutifulness	C3	−.08	<b>.57</b>	−.17	<b>.59</b>	−.06	−.09	<b>.52</b>	−.23	<b>.43</b>	−.13	−.02	<b>.50</b>	<b>.44</b>	.33	−.15
Achievement-striving	C4	.14	<b>.82</b>	−.14	.07	.11	.07	<b>.78</b>	−.05	.16	.04	.04	.22	<b>.75</b>	.20	.03
Self-discipline	C5	.07	<b>.78</b>	−.25	.13	−.12	.15	<b>.81</b>	−.19	.03	−.09	.19	.28	<b>.77</b>	.03	−.08
Cautiousness	C6	−.45	<b>.49</b>	−.32	.38	−.07	−.31	<b>.46</b>	−.48	.19	−.03	−.19	<b>.63</b>	.37	.01	−.01

Note. Loadings equal to or greater than .40 are in boldface type.

to all IPIP-NEO-120 items were ipsatized by subtracting each respondent's AI score from the item response and dividing by the standard deviation of AI. (Item numbers for the 29 pairs of opposite items and the SPSS syntax for generating ipsatized item scores are shown in the [Appendix A](#).)

30 Facet scores based on ipsatized item responses were computed, and these scores were correlated with the BFI and Mini-Marker acquaintance ratings in the Eugene-Springfield community sample to see if the correlations were higher than those based on scores where acquiescence was not controlled. In the Internet sample that completed the IPIP-NEO-120, 30 facet scores based on ipsatized responses were subjected to the same principal component analyses conducted earlier to see if a clearer factor structure emerged. The factor structures from both the normally-scored scales and acquiescence-controlled scales in this Internet sample were rotated in procrustean fashion toward the factor structure of the NEO PI-R (Costa & McCrae, 1992, p. 44) with an SPSS script developed by Fischer and Fontaine (2011) and made available by Wuensch (2014). The SPSS script also generated Tucker's coefficient of congruence, allowing an assessment factor structure similarity from each scoring method to the facet factor structure of the NEO PI-R.

AI scores ranged from 2.55 to 3.66 ( $SD = 3.04$ ) in the Eugene-Springfield sample. To see whether this overall acquiescence index was personally meaningful, AI was correlated with the BFI and Mini-Marker acquaintance ratings and with the NEO PI-R domain scores. A small but statistically significant and consistent negative correlation was found between AI and all three measures of Conscientiousness ( $r_s = -.15, -.15$ , and  $-.13$ , respectively; all  $p_s < .01$ ), indicating a slight tendency for individuals showing acquiescent responding to be unconscientious. This interpretation is reinforced by a correlation of  $-.16$  ( $p < .001$ ) between AI and Jackson's measure of individual profile reliability (Johnson, 2005) found in the Internet sample that completed the IPIP-NEO-300. The more important question, however, is whether correcting for acquiescence by ipsatizing item responses improves the validity of the 30 facet scale scores.

Consistent with the recommendations of Schriesheim and Hill (1981) against trying to counteract acquiescence, analyses with ipsatized scoring indicated that controlling for acquiescence decreases, rather than increases, the validity of the IPIP-NEO-120 facet scales. Whereas the normally-scored domain scores from the IPIP-NEO-120 correlated on average .46 with the BFI scales and .40 with the Mini-Marker scales in the community sample, domain scores based on ipsatized scoring correlated on average .08 with the BFI and .05 with the Mini-Markers. When scored normally, Tucker's coefficient of congruence between factors in the NEO PI-R and IPIP-NEO-120 were found to be .93, .97, .92, .87, and .95 for the N, E, O, A, and C factors in the Internet sample that completed the IPIP-NEO-120. When acquiescence was controlled with ipsatized scoring, the coefficients of congruence were .63, .81, .57, .81, and .67. These findings indicate that the factor structure of the IPIP-NEO-120 actually aligns very well with the factor structure of its parent, the NEO PI-R, and that there is no indication that imbalanced keying and acquiescence are problems for the IPIP-NEO-120.

#### 4. General discussion

The IPIP-NEO-120 has been freely distributed since its construction in October, 2000. Consequently, studies employing this instrument have begun to appear in the published literature. Research employing the IPIP-NEO-120 has addressed the following topics: behavior in online role playing games (Eladhari & Mateas, 2008;

Mosley, 2010), collaborative task performance (McGivney, Smeaton, & Lee, 2009), consumer choice in technology products (Saati, Salem, & Brinkman, 2005), cooperation in orthodontic patients (Gilbert, 2009), drinking in first-year college students (McAdams & Donnellan, 2009), genetics of myopia (van den Berg, Dirani, Chen, Haslam, & Baird, 2008), performance in financial day-traders (Lo, Repin, & Steenbarger, 2005), personality variability and interpersonal dysfunction (Clifton & Kuper, 2011), personnel selection (Li, Lai, & Kao, 2008), psychopathic personality traits (Miller & Lynam, 2012; Witt, Donnellan, & Blonigen, 2009), reactions to ostracism (McDonald & Donnellan, 2006), risk preference (Blackstone, Crabb, & Oswald, 2009), and traits of security personnel (Whalen & Gates, 2007). Until the publication of the current article, researchers have initiated their projects on the basis of informal communications from the author about preliminary, unpublished findings about the promising reliability and validity of the IPIP-NEO-120. Future researchers who are considering using the IPIP-NEO-120 now have a published source of information on its reliability and validity.

The construct validity of a psychological measure is, of course, never established by the publication of one study or even a large set of studies. Construct validation is a never-ending process involving continuous theory-testing with new studies (Hogan & Nicholson, 1988). The philosophy underlying the International Personality Item Pool project involves open-source sharing among professionals and encouraging revisions and experimentation with IPIP scales (Goldberg et al., 2006). I therefore expect others to experiment with the items in the International Personality Item Pool, looking for ways to measure the 30 facets that improve upon the length, content, reliability, validity, and factor structure of the IPIP-NEO-120.

A strength of the current study is its employment of large samples to identify the "best" four items (from an internal-consistency viewpoint) from each 10-item facet scale from the original IPIP-NEO and to verify that the IPIP-NEO-120 possesses acceptable reliability and validity for future research. My own sense from exhaustive examination of item-total correlations in each of the 30 facets is that researchers are unlikely to measure all 30 facets reliably with fewer than four items per scale. However, a limitation of the study is that alternative methods for constructing a shorter inventory of the 30 facets were not explored. Although the IPIP-NEO-120 is a *relatively* short inventory of the 30 facets, there may be ways of measuring the 30 facets more efficiently. For example, researchers who are willing to employ facet scales with unequal numbers of items may find that some facets can be reliably and validly measured with fewer than four items. Also, researchers who want to shorten administration time could develop more efficient computer adapted testing (CAT) versions of the IPIP-NEO using item response theory (e.g., Trippe & Harvey, 2003).

To encourage and facilitate improved measurement of the 30 facets of the FFM with short scales, future plans including placing item responses from the two large Internet samples used in the current study in the public domain for analyses by interested members of the personality research community. Hopefully this data sharing will encourage other IPIP users to share their data, helping the IPIP website to move closer toward its vision of serving as a "scientific collaboratory" that "may include raw data available for reanalysis" (International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits, n.d.). Science is meant to be an open, collaborative endeavor, limited only by the imagination of researchers. In that spirit, what the present study offers is not just the validation of the current IPIP-NEO-120 as the only end product of Internet

data analysis, but the underlying data set itself, grist for the mill of scientific imagination. Future research may use the data to construct a better 30-facet measure, but might also use these large data sets in imaginative, unforeseen ways.

## Acknowledgments

The author expresses his gratitude to the staff of the Oregon Research Institute for providing data from the Eugene-Springfield community sample. Portions of this research were presented at the 2nd Biennial Meeting of the Association for Research in Personality (Johnson, 2011).

## Appendix A

Computing the Content-Balanced Acquiescence Index and Ipsatizing the IPIP-NEO-120 Items.

Across the five factor domains the IPIP-NEO-120, 29 item pairs with opposite implications for personality (e.g., item 4, “Trust others” vs. item 94, “Distrust people”) were identified to compute an Acquiescence Index (AI) that was independent of item content. All pairs came from the same facet scale and are keyed in opposite directions. To verify that these opposite-content item pairs tended to be answered in opposite directions, Pearson correlations were computed for each item pair in the Internet sample that completed the IPIP-NEO-120. In every case, responses to the item pairs showed a significant negative correlation. Item numbers (see the column labeled “IPIP-120 Item No.” in Table 1 in the main text) for the 29 pairs are as follows: 4 and 94, 44 and 74, 59 and 119, 10 and 70, 15 and 105, 45 and 75, 20 and 110, 50 and 80, 25 and 115, 55 and 85, 2 and 62, 32 and 92, 7 and 97, 37 and 67, 12 and 102, 77 and 107, 36 and 96, 41 and 101, 16 and 106, 21 and 51, 56 and 116, 8 and 98, 38 and 68, 13 and 73, 43 and 103, 18 and 48, 23 and 113, 28 and 88, 58 and 118.

The SPSS syntax below computes AI for each individual as the mean of their responses to the 29 pairs of opposite-content items. The syntax then calculates the dispersion of their responses as the standard deviation of their responses to the items. Then, ipsatized item responses are calculated by subtraction AI from each of the 120 item responses and dividing by the dispersion of the opposite-content items.

```
COMPUTE acqavg = mean(I2,I4,I7,I8,I10,I12,I13,I15,I16,
I18,I20,I21,I23,I25,I28,I32,I36,I37,I38,I41,I43,I44,I45,
I48,I50,I51,I55,I56,I58,I59,I62,I67,I68,I70,I73,I74,I75,
I77,I80,I85,I88,I92,I94,I96,I97,I98,I101,I102,I103,
I105,I106,I107,I110,I113,I115,I116,I118,I119).
COMPUTE acqsd = sd(I2,I4,I7,I8,I10,I12,I13,I15,I16,
I18,I20,I21,I23,I25,I28,I32,I36,I37,I38,I41,I43,I44,
I45,I48,I50,I51,I55,I56,I58,I59,I62,I67,I68,I70,
I73,I74,I75,I77,I80,I85,I88,I92,I94,I96,I97,I98,I101,
I102,I103,I105,I106,I107,I110,
I113,I115,I116,I118,I119).
EXECUTE.
COMPUTE za1 = (I1-acqavg)/acqsd.
COMPUTE za2 = (I2-acqavg)/acqsd.
COMPUTE za3 = (I3-acqavg)/acqsd.
.
.
.
COMPUTE za120 = (I120-acqavg)/acqsd.
EXECUTE.
```

## References

- Adam, E. K., Doane, L. D., Zinbarg, R. E., Mineka, S., Craske, M. G., & Griffith, J. W. (2010). Prospective prediction of major depressive disorder from cortisol awakening responses in adolescence. *Psychoneuroendocrinology*, 35(6), 921–931. <http://dx.doi.org/10.1016/j.psyneuen.2009.12.007>.
- Ali, Z., Nisar, N., & Raza, A. (2011). Impact of personality traits on leadership styles of secondary school teachers. *Interdisciplinary Journal of Contemporary Research in Business*, 3(3), 610–617.
- Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: Macmillan.
- 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. [http://dx.doi.org/10.1016/0092-6566\(73\)90028-7](http://dx.doi.org/10.1016/0092-6566(73)90028-7).
- Blackstone, T. F., Crabb, J. C., & Oswald, F. L. (2009). *Risk preference elicitation and the role of personality and intelligence* (NPRST-TN-09-8). Millington, TN: Navy Personnel Research, Studies and Technology.
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54, 106–148. <http://dx.doi.org/10.1111/j.1467-6494.1986.tb00391.x>.
- Clifton, A., & Kuper, L. E. (2011). Self-reported personality variability across the social network is associated with interpersonal dysfunction. *Journal of Personality*, 79(2), 359–389. <http://dx.doi.org/10.1111/j.1467-6494.2010.00686.x>.
- Conway, J. M., Rogelberg, S. G., & Pitts, V. E. (2009). Workplace helping: Interactive effects of personality and momentary positive affect. *Human Performance*, 22(4), 321–339. <http://dx.doi.org/10.1080/08959280903120279>.
- 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.
- Craske, M. G., Waters, A. M., Nazarian, M., Mineka, S., Zinbarg, R. E., Griffith, J. W., et al. (2009). Does neuroticism in adolescents moderate contextual and explicit threat cue modulation of the startle reflex? *Biological Psychiatry*, 65(3), 220–226. <http://dx.doi.org/10.1016/j.biopsych.2008.07.020>.
- Edmonds, G. W., Goldberg, L. R., Hampson, S. E., & Barckley, M. (2013). Personality stability from childhood to midlife: Relating teachers' assessments in elementary school to observer- and self-ratings 40 years later. *Journal of Research in Personality*, 47, 505–513. <http://dx.doi.org/10.1016/j.jrp.2013.05.003>.
- Eladhari, M. P., & Mateas, M. (2008). *Semi-autonomous avatars in world of minds. ACE '08: Proceedings of the 2008 international conference on advances in computer entertainment technology*. New York: Association for Computing Machinery. <http://dx.doi.org/10.1145/1501750.1501798>, pp. 201–208.
- Fischer, R., & Fontaine, J. R. J. (2011). Methods for investigating structural equivalence. In D. Matsumoto & F. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 179–215). Cambridge University Press.
- Fletcher, T. D., & Nusbaum, D. N. (2008). Trait competitiveness as a composite variable: Linkages with facets of the big-five. *Personality and Individual Differences*, 45(4), 312–317. <http://dx.doi.org/10.1016/j.paid.2008.04.020>.
- Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. Thousand Oaks, CA: Sage.
- Gilbert, A. (2009). Patient personality and lingual orthodontics. *Orthodontic News*, June 30, 2009. <<http://www.orthodonticproductsonline.com/orp-orthodontic-news/13476-patient-personality-and-lingual-orthodontics-2009-07-02>> Retrieved.
- Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monograph*, 7, No. 72–2.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Goldberg, L. R. (2008). The Eugene-Springfield community sample: Information available from the research participants. *ORI Technical Report*, 48(1). March <[http://iipip.ori.org/iipip/iipip/iipip/iipip/ORI\\_TechnicalReport\\_ESCS\\_Mar08.pdf](http://iipip.ori.org/iipip/iipip/iipip/iipip/ORI_TechnicalReport_ESCS_Mar08.pdf)> Retrieved.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96. <http://dx.doi.org/10.1016/j.jrp.2005.08.007>.
- Goldberg, L. R. (n.d.-a). IPIP Scale-construction procedures. In *International Personality Item Pool: A scientific collaboratory for the development of advanced measures of personality traits and other individual differences*. <<http://iipip.ori.org/newScaleConstruction.htm>> Retrieved.
- Goldberg, L. R. (n.d.-b). Translating IPIP items. In *International Personality Item Pool: A scientific collaboratory for the development of advanced measures of personality traits and other individual differences*. <<http://iipip.ori.org/newItemTranslations.htm>> Retrieved.
- Gough, H. G. (1987). *CPI administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.
- Griffin, B., & Hesketh, B. (2004). Why openness to experience is not a good predictor of job performance. *International Journal of Selection and Assessment*, 12(3), 243–251. <http://dx.doi.org/10.1111/j.0965-075X.2004.278.1.x>.



- Hagger-Johnson, G., & Whiteman, M. C. (2007). Conscientiousness facets and health behaviors: A latent variable modeling approach. *Personality and Individual Differences*, 43(5), 1235–1245. <http://dx.doi.org/10.1016/j.paid.2007.03.014>.
- Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate data analysis* (6th ed.). New Jersey: Pearson Educational Inc.
- Hauner, K. K. Y., Adam, E. K., Mineka, S., Doane, L. D., DeSantis, A. S., Zinbarg, R., et al. (2008). Neuroticism and introversion are associated with salivary cortisol patterns in adolescents. *Psychoneuroendocrinology*, 33(10), 1344–1356. <http://dx.doi.org/10.1016/j.psyneuen.2008.07.011>.
- Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. *American Psychologist*, 43(8), 621–626. <http://dx.doi.org/10.1037/0003-066X.43.8.621>.
- International Personality Item Pool (n.d.). A scientific collaboratory for the development of advanced measures of personality traits and other individual differences. Internet web site <<http://ipip.ori.org/>>.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). New York, NY: Guilford Press.
- Johnson, J. A. (2000). *Web-based personality assessment*. Paper presented at the 71st annual meeting of the eastern psychological association, Baltimore, MD.
- Johnson, J. A. (2001, May). *Screening massively large data sets for non-responsiveness in web-based personality inventories*. Invited talk to the joint Bielefeld-Groningen Personality Research Group, University of Groningen, The Netherlands. <<http://www.personal.psu.edu/~j5j/papers/ConferencePapers/2001BGRPG.pdf>> Retrieved.
- Johnson, J. A. (2005). Ascertaining the validity of web-based personality inventories. *Journal of Research in Personality*, 39, 103–129. <http://dx.doi.org/10.1016/j.jrp.2004.09.009>.
- Johnson, J. A. (2009, July). *Calibrating personality self-report scores to acquaintance ratings*. Poster presented at the first stand-alone conference of the Association for Research in Personality, Evanston, IL.
- Johnson, J. A. (2011, June). *Development of a short form of the IPIP-NEO Personality Inventory*. Poster presented at the 2nd Biennial Meeting of the Association for Research in Personality, Riverside, CA.
- Leininger, A. (2002). Vietnamese-American personality and acculturation: An exploration of relations between personality traits and cultural goals. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor model of personality across cultures* (pp. 197–225). New York: Kluwer Academic/Plenum Publishers.
- Lewis, Alison R., Zinbarg, Richard E., Mineka, Susan, Craske, Michelle G., Epstein, Alyssa, & Griffith, James W. (2010). The relationship between anxiety sensitivity and latent symptoms of emotional problems: A structural equation modeling approach. *Behaviour Research and Therapy*, 48, 761–769. <http://dx.doi.org/10.1016/j.brat.2010.05.001>.
- Li, Y. -M., Lai, C. -Y., & Kao, C. -P. (2008). Incorporate personality trait with support vector machine to acquire quality matching of personnel recruitment. In *4th Proc. international conference on business and information 2008 (BAI 2008)* (pp. 1–11). CD-ROM:B2-315.
- Lo, A. W., Repin, D. V., & Steenbarger, B. N. (2005). Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review*, 95(2), 352–359. <http://dx.doi.org/10.2139/ssrn.690501>.
- McAdams, K. K., & Donnellan, M. B. (2009). Facets of personality and drinking in first-year college students. *Personality and Individual Differences*, 46, 207–212. <http://dx.doi.org/10.1016/j.paid.2008.09.028>.
- McDonald, M. M., & Donnellan, M. B. (2006). Is ostracism a strong situation? The influence of personality in reactions to rejection. *Journal of Research in Personality*, 40, 614–618. <http://dx.doi.org/10.1016/j.jrp.2012.05.008>.
- McGivney, S., Smeaton, A. F., & Lee, H. (2009). The effect of personality on collaborative task performance and interaction. In E. Bertino & J. B. D. Joshi (Eds.), *Collaborative computing: Networking, applications and worksharing: Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering volume 10* (pp. 499–511). Berlin: Springer. [http://dx.doi.org/10.1007/978-3-642-03354-4\\_38](http://dx.doi.org/10.1007/978-3-642-03354-4_38).
- Meehl, P. E. (1970). Some methodological reflections on the difficulties of psychoanalytic research. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science. Analyses of theories and methods of physics and psychology* (Vol. IV, pp. 403–416). Minneapolis: University of Minnesota Press.
- Miller, J. D., & Lynam, D. R. (2012). An examination of the psychopathic personality inventory's nomological network: A meta-analytic review. *Personality Disorders: Theory, Research, and Treatment*, 3(3), 305–326. <http://dx.doi.org/10.1037/a0024567>.
- Morizot Ainsworth & Reise (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 407–423). New York: Guilford Press.
- Mosley, I. D. (2010). Personality and character selection in World of Warcraft. *McNair Scholars Research Journal*, 6 (Iss. 1, Article 12) <[http://scholarworks.boisestate.edu/mcnair\\_journal/vol6/iss1/12](http://scholarworks.boisestate.edu/mcnair_journal/vol6/iss1/12)> Retrieved.
- Möttus, R., Pullmann, H., & Allik, J. (2006). Toward more readable Big Five personality inventories. *European Journal of Psychological Assessment*, 22(3), 149–157. <http://dx.doi.org/10.1027/1015-5759.22.3.149>.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Ogot, M., & Okudan, G. E. (2006). The Five-Factor Model personality assessment for improved student design team performance. *European Journal of Engineering Education*, 31(5), 517–529. <http://dx.doi.org/10.1080/03043790600797335>.
- Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, 44, 53–61. <http://dx.doi.org/10.1016/j.jrp.2009.10.005>.
- Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers & R. D. Hays (Eds.), *Assessing quality of life in clinical trials: Methods of practice* (2nd Ed., pp. 55–73). Oxford: Oxford University Press.
- Rorer, L. G. (1965). The great response style myth. *Psychological Bulletin*, 63, 129–156.
- Saati, B., Salem, M., & Brinkman, W. -P. (2005). Towards customized user interface skins: investigating user personality and skin colour. In *Proceedings of HCI 2005* (Vol. 2, pp. 89–93). <[http://www.academia.edu/341112/Towards\\_Customized\\_User\\_Interface\\_Skins\\_Investigating\\_User\\_Personality\\_and\\_Skin\\_Colour](http://www.academia.edu/341112/Towards_Customized_User_Interface_Skins_Investigating_User_Personality_and_Skin_Colour)> Retrieved.
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63, 506–516. [http://dx.doi.org/10.1207/s15327752jpa6303\\_8](http://dx.doi.org/10.1207/s15327752jpa6303_8).
- Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, 41, 1101–1114. <http://dx.doi.org/10.1177/001316448104100420>.
- Sibley, C. G. (2012). The Mini-IPIP6: Item Response Theory analysis of a short measure of the big-six factors of personality in New Zealand. *New Zealand Journal of Psychology*, 41, 21–31.
- Sibley, C. G., Luyten, N., Purnomo, M., Moberly, A., Wootton, L. W., Hammond, M. D., et al. (2011). The Mini-IPIP6: Validation and extension of a short measure of the Big-Six factors of personality in New Zealand. *New Zealand Journal of Psychology*, 40, 142–159.
- Siddiqui, K. (2011). Personality influences mobile phone usage. *Interdisciplinary Journal of Contemporary Research in Business*, 3(3), 554–565.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, 94, 718–737. <http://dx.doi.org/10.1037/0022-3514.94.4.718>.
- SPSS, Inc. (1999). *SPSS Base 10.0 applications guide*. Chicago, Author.
- Sutton, J. M., Mineka, S., Zinbarg, R. E., Craske, M. G., Griffith, J. W., Rose, R. D., et al. (2011). The relationships of personality and cognitive styles with self-reported symptoms of depression and anxiety. *Cognitive Therapy and Research*, 35(4), 381–393. <http://dx.doi.org/10.1007/s10608-010-9336-9>.
- Tripe, D. M., & Harvey, R. J. (2003, April). *Item response theory analysis of the IPIP big-five scales*. Paper presented at the 18th annual conference of the society for industrial and organizational psychology, Orlando, FL.
- U.S. Department of Labor Employment and Training Administration. (1999). *Testing and assessment: An employer's guide to good practices*. <<http://uniformguidelines.com/testassess.pdf>> Retrieved.
- van den Berg, R., Dirani, M., Chen, C. Y., Haslam & Baird, P. N. (2008). Myopia and personality: The genes in myopia (GEM) personality study. *Investigative Ophthalmology and Visual Science*, 49(3), 882–886. <http://dx.doi.org/10.1167/iovs.07-0930>.
- Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98, 281–300. <http://dx.doi.org/10.1037/a0017908>.
- Wallace, C., & Chen, G. (2006). A multilevel integration of personality, climate, self-regulation, and performance. *Personnel Psychology*, 59, 529–557. <http://dx.doi.org/10.1111/j.1744-6570.2006.00046.x>.
- Whalen, T., & Gates, C. (2007). A psychological profile of defender personality traits. *Journal of Computers*, 2(2), 84–93. <http://dx.doi.org/10.4304/jcp.2.2.84-93>.
- Witt, E. A., Donnellan, M. B., & Blonigen, D. M. (2009). Using existing self-report inventories to measure the psychopathic personality traits of Fearless Dominance and Impulsive Antisociality. *Journal of Research in Personality*, 43, 1006–1016. <http://dx.doi.org/10.1016/j.jrp.2009.06.010>.
- Wolfe, R. N. (1993). A commonsense approach to personality measurement. In K. H. Craik, R. Hogan, & R. N. Wolfe (Eds.), *Fifty years of personality psychology* (pp. 269–290). New York: Plenum.
- Wuenschk, K. L. (2014). Comparing two groups' factor structures. In *SPSS Programs (Syntax)*. <<http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Programs.htm>> Retrieved.