

Development of a Short Form of the IPIP-NEO Personality Inventory Poster #1.18

Presented at the 2nd Biennial ARP Conference
Riverside, CA June 17, 2011

John A. Johnson
Pennsylvania State University, DuBois



Abstract

➤ One of the first personality measures to be created from the International Personality Item Pool (IPIP; Goldberg, et al., 2006) was a 300-item inventory (Goldberg, 1999) designed to measure constructs similar to those in the NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1985). The IPIP version of the NEO PI-R has been dubbed the IPIP-NEO. One disadvantage of the IPIP-NEO is that it is even longer than the original 240-item NEO PI-R. This poster describes the development and validation of a more usable 120-item version of the IPIP-NEO. Analyses with Goldberg's (2008) Eugene-Springfield community sample (N=501), two large Internet samples (Ns = 307,313 and 619,150) and a local sample (N=160) indicate that the psychometric properties of the 120-item IPIP-NEO compare favorably to the properties of the longer form. Although short, public-domain measures of the five major personality factors abound, only the 120-item IPIP-NEO has demonstrated psychometrically acceptable scales for the 30 NEO PI-R facets.

Background

➤ Copyright restrictions on commercial personality inventories limit certain kinds of research, including collecting data on the World Wide Web. The International Personality Item Pool (IPIP; Goldberg, et al., 2006), a repository of public-domain personality items, was created to overcome the limitations of commercial inventories. 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 in the NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1985). The original IPIP version of the NEO PI-R has been dubbed the IPIP-NEO-300. One disadvantage of the IPIP-NEO-300 is that it is even longer than the original 240-item NEO PI-R. An unpublished paper by Johnson (2001) describes the creation of a Web-based version of the IPIP-NEO-300 and the initial development of a 120-item subset version, or IPIP-NEO-120. This poster describes the further validation of the IPIP-NEO-120.

Scale Development Procedures

➤ Data from an Internet sample of over 20,000 respondents (Johnson, 2005) was first screened to eliminate duplicate and invalid protocols. Scale development then proceeded in three phases. The first phase iteratively removed items with the lowest item-total correlations for each of the 10 item-facet scales until four items remained. Coefficient alphas for the 10-item facet scales had 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 C3 Dutifulness, which was .69.

➤ The second phase involved examining the item content of every 4-item scale for three properties: near-duplicate items, references to disabilities or other areas that may result in legal problems, and fidelity to the content of items on Costa and McCrae's original NEO PI-R. Any item whose wording I judged to be too close to one of the other three items was replaced with a substitute that maintained the highest level of alpha reliability. For example, on O4 Adventurousness, two finalists for the four-item scale were "Dislike changes" and "Don't like the idea of change." The latter was replaced with "Prefer variety to routine."

➤ The third phase of scale development was ascertaining that alphas for the five domain scales 30 facet scales were acceptable for both males and females. All alphas were $\geq .70$ except E4 Activity Level (.69), O3 Emotionality (.68), and O6 Liberalism (.64) for males and O3 Emotionality (.66), O6 Liberalism (.63), A6 Sympathy (.67), and C3 Dutifulness (.69) for females.

New Analyses

➤ Four samples were used to further assess the psychometric properties of the IPIP-NEO-120.

➤ Goldberg's (2008) Eugene-Springfield community sample (N=481) had completed the IPIP-NEO-300 and Costa and McCrae's NEO PI-R; 420 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 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 10 years, more than 300,000 persons have 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 site is to educate respondents about the five-factor model of personality; respondents receive narrative feedback upon completing either inventory. Participation is anonymous and recorded item response data are used primarily for norming and psychometric evaluation of the scales. 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 third sample was a set of 160 participants who completed the IPIP-NEO-300 and were judged by acquaintances on descriptions of the five domain 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 primary evidence of the relative validity of the long and short scales.

Results and Discussion

➤ Results of the analyses data from the four samples are displayed in Table 1. Because the IPIP-NEO-300 and IPIP-NEO-120 are meant to reliably represent the constructs in the original NEO PI-R, the primary reliability and validity data of interest are contained in the first four columns of Table 1. Coefficient alphas ranged from .71 to .89 for the facet scales (mean $\alpha = .80$) and from .90 to .94 for the domain scales of the IPIP-NEO-300; facet alphas for the IPIP-NEO-120 ranged from .47 to .80 (mean $\alpha = .68$) and from .81 to .88 for the domain scales. The alphas for the 4-item scales of the IPIP-NEO-120 are, as expected, lower than the alphas for the 10-item scales. Nonetheless, these values indicate that the scales are reliable enough for research purposes. The scales of the two instruments show similar validity in the form of correlations with the original NEO PI-R scales, with an average correlations of .73 (.94 corrected for attenuation due unreliability) for the long scales and .66 (.91 corrected for attenuation) for the short scales. Although the IPIP-NEO scales do not measure exactly the same constructs as the NEO PI-R, they measure highly similar constructs.

➤ Correlations between the long and short IPIP-NEO scales and acquaintance ratings of the five major personality factors as assessed by the BFI and Mini-Markers are shown in the next two columns of Table 1. The correlations were expected to be lower because of method differences and the different operationalizations of the five factors. Nonetheless, the BFI validity coefficients averaged .36 and .32 and the Mini-Marker correlations, .31 and .28, for the long and short IPIP-NEO scales, respectively.

➤ The Internet samples allowed a check on alpha reliabilities to see if the stand-alone IPIP-NEO-120 scale reliabilities differed from short scales scored within the IPIP-NEO-300. They did not. Average alpha for the long scales was .84 and .77 for the short scales, whether embedded in the long form or scored as a separate 120-item measure. The Internet samples also allowed an investigation of factor structure. Full results are too complex to present in this poster. Briefly, most—but not all—of the scales loaded highly on the expected factor. Researchers desiring a factorially pure FFM inventory may be dissatisfied with the factor structures of IPIP-NEO measures.

➤ If a researcher's concern is more with reliability and external validity, they will be encouraged by the correlations with acquaintance ratings of the domain and facet constructs, which averaged .37 for the long scales and .34 for the short scales.

➤ In conclusion, the IPIP-NEO-120 appears to be a reasonably reliable and valid public-domain measure of the five personality factors at the facet level.

References

- Goldberg, L. R. (1999a). 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, March). The Eugene-Springfield community sample: Information available from the research participants. *ORI Technical Report*, 48(1).
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96. doi:10.1016/j.jrp.2005.08.007
- 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. (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/screening.html>
- Johnson, J. A. (2005). Ascertaining the validity of web-based personality inventories. *Journal of Research in Personality*, 39, 103-129. doi: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.
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63, 506-516. doi: 10.1207/s15327752jpa6303_8

Development of a Short Form of the IPIP-NEO Personality Inventory Poster #1.18

John A. Johnson
Pennsylvania State University, DuBois



Table 1

SCALE LABELS	NEO PI-R	Eugene-Springfield Sample (Ns = 420-481)								Internet Sample (N=307,313)		Internet Sample (N=619,150)		Internet Sample (N=307,313)		Internet Sample (N=619,150)		Local Sample (N=160)	
		Alpha	Correlations with NEO PI-R		Correlations with BFI Informant Judgments		Correlations with MiniMarker Informant Judgments		Alpha	Alpha	Loading on Expected Factor		Loading on Expected Factor		Correlation with Informant Rating				
IPIP-NEO		IPIP 300	IPIP 120	IPIP 300	IPIP 120	IPIP 300	IPIP 120	IPIP 300	IPIP 120	IPIP 300	IPIP 120	IPIP 120	IPIP 300	IPIP 120	IPIP 120	IPIP 300	IPIP 120		
<i>Neuroticism</i>	N	.94	.88	.88	.87	.47	.46	.34	.33	.95	.90	.90					.52	.49	
ANXIETY	N1	.83	.71	.76	.76	.45	.44	.32	.30	.86	.78	.78	.86	.76	.54	.44	.40		
ANGER	N2	.88	.77	.77	.71	.41	.32	.36	.29	.91	.86	.87	.74	.73	.72	.55	.55		
DEPRESSION	N3	.89	.80	.81	.76	.39	.38	.30	.28	.91	.86	.85	.69	.59	.40	.61	.61		
SELF-CONSCIOUSNESS	N4	.80	.63	.73	.60	.25	.18	.10	.06	.82	.72	.70	.51	.23	.00	.47	.33		
IMMODERATION	N5	.77	.69	.74	.65	.24	.25	.24	.21	.77	.71	.69	.53	.44	.52	.33	.33		
VULNERABILITY	N6	.82	.70	.78	.74	.36	.36	.22	.23	.85	.76	.76	.82	.70	.52	.44	.43		
<i>Extraversion</i>	E	.92	.84	.89	.85	.51	.49	.48	.45	.94	.89	.89				.43	.42		
FRIENDLINESS	E1	.87	.77	.76	.68	.43	.42	.42	.41	.88	.81	.81	.82	.81	.82	.41	.39		
GREGARIOUSNESS	E2	.79	.60	.78	.73	.36	.37	.33	.35	.88	.79	.79	.85	.83	.79	.42	.37		
ASSERTIVENESS	E3	.84	.75	.81	.73	.45	.29	.46	.30	.85	.83	.85	.59	.45	.37	.47	.39		
ACTIVITY LEVEL	E4	.71	.68	.72	.63	.34	.31	.31	.27	.71	.70	.69	.28	.32	.22	.37	.36		
EXCITEMENT-SEEKING	E5	.77	.67	.67	.59	.15	.15	.12	.10	.84	.75	.73	.65	.63	.53	.46	.43		
CHEERFULNESS	E6	.81	.71	.77	.69	.34	.28	.29	.24	.82	.79	.79	.73	.70	.75	.39	.42		
<i>Openness to Experience</i>	O	.92	.85	.87	.84	.58	.57	.52	.49	.90	.82	.81				.30	.27		
IMAGINATION	O1	.82	.70	.74	.69	.37	.29	.34	.29	.84	.75	.74	.68	.58	.54	.26	.20		
ARTISTIC INTERESTS	O2	.85	.72	.80	.76	.45	.45	.37	.34	.80	.74	.74	.63	.71	.69	.36	.36		
EMOTIONALITY	O3	.81	.67	.71	.65	.34	.33	.33	.25	.77	.66	.65	.48	.31	.23	.42	.39		
ADVENTUROUSNESS	O4	.77	.66	.72	.62	.40	.38	.28	.30	.80	.70	.70	.53	.54	.56	.28	.19		
INTELLECT	O5	.86	.78	.81	.75	.43	.41	.31	.40	.84	.74	.73	.79	.78	.78	.24	.23		
LIBERALISM	O6	.86	.76	.71	.63	.40	.36	.44	.33	.77	.64	.63	.55	.47	.45	.35	.34		
<i>Agreeableness</i>	A	.90	.81	.83	.76	.38	.33	.35	.31	.92	.85	.86				.29	.25		
TRUST	A1	.82	.70	.78	.73	.29	.31	.26	.27	.88	.86	.85	.53	.48	.35	.30	.28		
MORALITY	A2	.74	.62	.65	.54	.22	.15	.22	.16	.78	.74	.74	.77	.65	.50	.14	.09		
ALTRUISM	A3	.77	.65	.68	.54	.32	.23	.28	.19	.82	.74	.73	.70	.73	.77	.16	.14		
COOPERATION	A4	.72	.56	.72	.62	.40	.31	.35	.29	.77	.70	.71	.81	.64	.41	.25	.25		
MODESTY	A5	.76	.63	.71	.64	.10	.08	.08	.07	.77	.75	.73	.49	.48	.46	.22	.24		
SYMPATHY	A6	.75	.68	.62	.55	.21	.18	.23	.21	.77	.70	.72	.70	.67	.71	.44	.35		
<i>Conscientiousness</i>	C	.92	.84	.84	.80	.46	.45	.45	.44	.94	.90	.90				.42	.42		
SELF-EFFICACY	C1	.79	.57	.68	.59	.22	.25	.21	.24	.81	.63	.77	.68	.72	.74	.26	.25		
ORDERLINESS	C2	.83	.76	.77	.68	.45	.41	.49	.43	.85	.82	.83	.66	.58	.52	.57	.52		
DUTIFULNESS	C3	.71	.47	.60	.53	.32	.27	.27	.24	.78	.67	.67	.57	.52	.44	.35	.33		
ACHIEVEMENT-STRIVING	C4	.79	.68	.71	.57	.32	.35	.30	.33	.82	.79	.79	.82	.78	.75	.37	.41		
SELF-DISCIPLINE	C5	.85	.66	.77	.72	.31	.31	.30	.31	.89	.73	.71	.78	.81	.77	.37	.32		
CAUTIOUSNESS	C6	.76	.70	.69	.61	.32	.21	.28	.19	.84	.86	.88	.49	.46	.37	.28	.27		