Psychological Assessment
2010, Vol. 22, No. 1, 43–49

1040-3590/10/$12.00 DOI: 10.1037/a0017371

Confirmatory Factor Analysis of the M5-50: An Implementation of the International Personality Item Pool Item Set

Alan Socha, Christopher A. Cooper, and David M. McCord
Western Carolina University

Goldberg's International Personality Item Pool (IPIP; Goldberg, 1999) provides researchers with public-domain, free-access personality measurement scales that are proxies of well-established published scales. One of the more commonly used IPIP sets employs 50 items to measure the 5 broad domains of the 5-factor model, with 10 items per factor. The M5-50 (McCord, 2002) is a specific ordering and presentation of this 50-item set. Using data from a sample of 760 faculty, staff, and students at a midsized university, the authors assessed the reliability and construct validity of the M5-50. Cronbach's alphas on the 5 scales ranged from acceptable to excellent. Confirmatory factor analysis indicated reasonably good model fit. Researchers who wish to measure personality would be well advised to consider using the M5-50.

Keywords: IPIP, five-factor model of personality, confirmatory factor analysis, M5-50

Goldberg's International Personality Item Pool (IPIP; Goldberg, 1999) was initiated in 1996 with the goal of circumventing the severe constraints on personality research imposed by the commercialization and copyrighting of most major personality assessment instruments. Goldberg's intent was to develop a repository of personality items that were contextualized and therefore longer than single-word traits adjectives, shorter than most existing personality questionnaire items, and readily translatable into multiple languages. The original pool of 1,252 has now grown to 2,413 similarly formatted self-report personality items.

With a large adult sample, these items have been correlated with numerous scales from existing published personality tests, questionnaires, and marker sets, including the NEO Personality Inventory—Revised (NEO-PI–R; Costa & McCrae, 1992), the California Psychological Inventory (CPI; Gough & Bradley, 1996), the Temperament and Character Inventory (TCI; Cloninger, 1994), the Hogan Personality Inventory (HPI; Hogan & Hogan, 1992), and the Sixteen Personality Factor Scale (16PF; Conn & Rieke, 1994). By selecting IPIP items most highly correlated with the original scales, "proxy" scales may be developed that exhibit psychometric properties very similar to those of the parent scales. The IPIP thus allows free access to reliable and valid personality assessment instruments for anyone interested in personality research and is comparable in many respects to open-source software.

Alan Socha, Institutional Planning and Effectiveness, Western Carolina University; Christopher A. Cooper, Public Policy Institute and the Department of Political Science and Public Affairs, Western Carolina University; David M. McCord, Department of Psychology, Western Carolina University.

The data analyzed here were originally collected by Christopher A. Cooper, Lauren Golden, and Alan Socha in a study designed to explore relationships between personality traits and political opinions. Their manuscript is currently under review elsewhere. We took advantage of the large and diverse participant sample to conduct detailed psychometric analyses of the M5-50.

Correspondence concerning this article should be addressed to David M. McCord, Department of Psychology, Western Carolina University, Cullowhee, NC 28723. E-mail: mccord@wcu.edu

The popularity of the IPIP items and scales has increased rapidly. Items have been translated into more then 25 languages. In a 2006 publication, Goldberg et al. noted that more than 80 IPIP-related publications were listed on the IPIP website. At the time of this writing the number had grown to almost 350. Goldberg et al. attributed the remarkable success of this project to several factors: use of the IPIP is cost-free; all items are readily visible and retrievable via the Internet; and there are no copyright restrictions—items may be used in any order, interspersed with other items, administered via the web, modified, and translated, with no permission required.

The IPIP website (Goldberg, 1999) provides internal consistency values and scale-level correlations with the original instruments. Not surprisingly, because of the method of scale construction (see Goldberg et al., 2006), internal consistency values for the IPIP proxy scales tend to be as high as, or somewhat higher than, values for the parent scales. Similarly, correlations between the IPIP scales and the parent scales are generally quite high. Even so, one cannot assume that the proxy scales are measuring the same construct as the parent scale. For example, one extraversion measure may exhibit good internal consistency and correlate strongly with another extraversion measure, yet it may still have notably different conceptual content. Thus, although a high correlation with the parent scale supports the validity of an IPIP proxy scale, it is still important to establish the psychometric properties of the IPIP scale itself, independent of the parent scale. This is particularly true with regard to validity. Without this step, researchers cannot be certain that the IPIP proxy scale measures what it purports to measure.

One popular set of IPIP items includes the 50 that were selected to construct proxies for the broad domain scores of Costa and McCrae's (1992) NEO-PI–R: Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to Experience (O). The 10 items that most strongly correlate with the each of the five broad domain scales from the NEO-PI–R were identified. For example, the 10 items from the full IPIP item set with highest correlations with the E scale of the NEO-PI–R were

Table 1
Confirmatory Factor Analysis Models

Item #	Coding direction	Text	Proposed model	E model	A model	C model	N model	O model
1	+	Have a vivid imagination	O	O	O	C	N	O
2	+	Believe in the importance of art	O	O	O	O	O	O
3	−	Seldom feel blue	N	N	N	N	N	N
4	−	Have a sharp tongue	A	E	A	A	N	O
5	−	Am not interested in abstract ideas	O	O	O	C	N	O
6	−	Find it difficult to get down to work	C	E	C	C	N	C
7	+	Panic easily	N	E	N	C	N	N
8	+	Tend to vote for liberal political candidates	O	E	A	C	O	O
9	−	Am not easily bothered by things	N	E	A	C	N	N
10	+	Make friends easily	E	E	A	E	E	O
11	+	Often feel blue	N	N	N	N	N	O
12	+	Get chores done right away	C	E	C	C	C	C
13	−	Suspect hidden motives in others	A	A	A	A	N	A
14	−	Rarely get irritated	N	N	A	N	N	N
15	−	Do not like art	O	O	O	O	N	O
16	+	Dislike myself	N	N	N	N	N	N
17	−	Keep in the background	E	E	E	E	E	O
18	−	Do just enough work to get by	C	C	C	C	N	C
19	+	Am always prepared	C	C	C	C	N	C
20	−	Tend to vote for conservative political candidates	O	E	O	C	N	O
21	−	Feel comfortable with myself	N	E	A	N	N	O
22	−	Avoid philosophical discussions	O	O	O	O	N	O
23	−	Waste my time	C	C	C	C	C	C
24	+	Believe that others have good intentions	A	A	A	C	N	A
25	−	Am very pleased with myself	N	E	A	C	N	O
26	−	Have little to say	E	E	E	E	N	E
27	+	Feel comfortable around other people	E	E	E	E	N	O
28	+	Am often down in the dumps	N	N	N	N	N	N
29	−	Do not enjoy going to art museums	O	O	O	O	N	O
30	+	Have frequent mood swings	N	N	N	N	N	O
31	−	Don't like to draw attention to myself	E	E	E	E	N	E
32	−	Insult people	A	A	A	C	N	A
33	+	Have a good word for everyone	A	E	A	C	A	A
34	−	Get back at others	A	A	A	A	N	A
35	+	Carry out my plans	C	C	C	C	C	C
36	−	Would describe my experiences as somewhat dull	E	E	E	C	N	O
37	+	Carry the conversation to a higher level	O	E	O	C	O	O
38	−	Don't see things through	C	C	C	C	N	C
39	+	Am skilled in handling social situations	E	E	E	E	E	E
40	+	Respect others	A	E	A	A	A	A
41	+	Pay attention to details	C	C	C	C	N	C
42	+	Am the life of the party	E	E	E	E	E	E
43	+	Enjoy hearing new ideas	O	O	A	O	O	O
44	+	Accept people as they are	A	E	A	A	A	O
45	−	Don't talk a lot	E	E	E	E	N	E
46	−	Cut others to pieces	A	A	A	C	N	A
47	+	Make plans and stick to them	C	C	C	C	C	C
48	+	Know how to captivate people	E	E	E	E	E	E
49	+	Make people feel at ease	A	E	A	C	A	O
50	−	Shirk my duties	C	C	C	C	C	C

Note. Coding direction: + means that responses are coded in a positive direction; − means that responses are reverse-coded. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; O = Openness to Experience.

selected as the IPIP version of the E scale, and so forth, for the five factors.[1] The M5-50 (McCord, 2002) is a freely available specific ordering of these 50 items, with a cover page, instructions, and spaces for demographic and identifying data.

Results of previous examinations of the psychometric properties of this 50-item set have been mixed. In general, scale reliability estimates have tended to be quite high (see, e.g., Baird, Le, & Lucas, 2006; also see International Personality Item Pool, http://ipip.ori.org). On the other hand, results of confirmatory factor analyses have been less positive. For example, using a convenience sample of 353 undergraduate students, Lim and Ployhart (2006) fit each personality factor at the item level and found an average root-mean-square error of approximation (RMSEA) of .12, comparative fit index (CFI) of .83, and standardized root-

[1] See Goldberg et al. (2006) for a more complete description of the procedures used for IPIP scale development.

mean-square residual (SRMR) of .07. Although these results are suggestive, the authors noted that the results should not be treated as definitive given the relatively small and unrepresentative sample. It should be noted that negative confirmatory factor analysis (CFA) results have also been obtained with other five-factor model instruments (Borkenau & Ostendorf, 1990; Church & Burke, 1994; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). Indeed, due in part to construct overlap within the five-factor model itself, Donnellan, Oswald, Baird, and Lucas (2006, p. 197) suggested that "it might not be possible to obtain reasonable fit from a CFA perspective on many or even most Big Five inventories." The purpose of the present study was to provide support for the reliability and construct validity of the M5-50 (and thus this specific 50-item IPIP set). We expanded on these previous studies by obtaining a larger, more diverse sample and by offering a more robust CFA methodology with which we attempted to rule out alternative models.

Method

Sample

To establish the reliability and construct validity of the M5-50, we distributed a web-based questionnaire to the universe of faculty, staff, and students at a midsized public university in the southeast United States. A total of 760 volunteers responded, consisting of 451 students (120 men and 331 women), 145 faculty members (66 men and 79 women), and 164 university staff (57 men and 107 women). Approximately 50% of the sample had less than a college degree, 68% were women, and 90% were White.

This sampling strategy is notable for two reasons. First, existing studies using the IPIP item set frequently have relied on student samples. For example, Lim and Ployhart (2006) assessed the validity of the IPIP item set, but they cautioned against overgeneralizing their findings because their sample consisted of only undergraduate students. To combat this "college sophomore problem" (Sears, 1986), we sampled students, faculty, and staff. According to Kam, Wilking, and Zechmeister (2007), university staff are similar to the population at large on a host of demographic and psychological indicators, helping to make our results more generalizable.

The second reason our sampling strategy is particularly useful is that most previous studies assessing the properties of IPIP scales have included relatively small samples. For example, Donnellan et al.'s (2006) work contains a sample size of just over 200. As we noted earlier, however, through our unique sampling strategy, we were able to accumulate data on over 750 individuals. We expect that this difference will be consequential. McCrae et al. (1996) noted the importance of sample size, stating that "increasing sample size is likely to give increasingly precise estimates of the population factor structure" (p. 563). Together, these two departures from previous research should help us produce factor estimates that are valid and reliable.

Materials

The M5-50 consists of the 50 items from the IPIP item set that measure the five broad domains of the NEO-PI–R. The process used to develop the IPIP proxy scales combines empirical, rational, and psychometric methods and is described in Goldberg et al. (2006). Briefly, for a given parent scale, such as the Extraversion broad domain of the NEO-PI–R, the 10 items from the IPIP with the highest correlations with the parent scale constitute the proxy. To minimize the potential distortion of a directional response set (acquiescence or nay-saying), approximately half of the items are worded negatively and reverse scored. Standard IPIP instructions were presented to participants, who responded on a 5-point Likert-type scale ranging from 1 (*inaccurate*) to 5 (*accurate*) with a neutral midpoint.

Procedure

Once the data were collected, we conducted a CFA, using EQS (Version 6.1), to determine if the M5-50 items fit best on their proposed factors for E, A, C, N, and O. To properly assess construct validity, we wanted to compare the model to potential alternative models. Unfortunately, including every possible model would require us to run 5^{50} different models. We did not choose this approach for two reasons. First, the sheer number of potential models made this approach impractical. Second, and most important, the vast majority of these models would have very little theoretical justification, no face validity, and much less construct validity. As a result, we chose a more theoretically grounded and fruitful method for selecting alternative models.

To pick these alternative models, we asked an expert panel of six clinical psychology graduate students to place each item in the M5-50 into the factor(s) where they believed the item best fit (they could choose up to three factors for each item). These responses were used to construct an alternative model for each personality factor. For example, we constructed the potential E model by taking all M5-50 items that were interpreted by at least one expert as fitting in E and assigning them there, keeping the loadings for

Table 2
Goodness-of-Fit Indices

Model	χ^2	*df*	*p*	CFI	RMSEA	SRMR
Proposed model	5291.169	1165	.000	.706	.068	.083
Extraversion model	6904.323	1167	.000	.591	.080	.112
Agreeableness model	5918.156	1165	.000	.661	.073	.092
Conscientiousness model	6938.607	1165	.000	.588	.081	.107
Neuroticism model	8854.192	1165	.000	.452	.093	.115
Openness to Experience model	7221.305	1165	.000	.568	.083	.097

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

all other items in their proposed factor. In this model, at least one member of our panel thought that Item 13 could be associated with E, so Item 13's loading was assigned to E. No panel members thought that Item 1 fit under E, so Item 1 was left loading in O, its proposed factor. Table 1 contains the potential models that our expert panel identified.

Table 3
Pattern and Structure Coefficients for Proposed Model

Item #	Pattern coefficient	Structure coefficient
	Agreeableness	
4	0.563	0.438
13	0.541	0.470
24	0.376	0.451
32	0.661	0.666
33	0.478	0.525
34	0.564	0.502
40	0.334	0.587
44	0.318	0.438
46	0.448	0.569
49	0.320	0.417
	Conscientiousness	
6	0.886	0.675
12	0.711	0.599
18	0.636	0.572
19	0.510	0.511
23	0.772	0.644
35	0.489	0.668
38	0.640	0.726
41	0.399	0.480
47	0.584	0.627
50	0.527	0.579
	Extraversion	
10	0.700	0.631
17	0.840	0.673
26	0.636	0.602
27	0.658	0.643
31	0.663	0.553
36	0.472	0.433
39	0.708	0.679
42	0.815	0.682
45	0.872	0.685
48	0.661	0.651
	Neuroticism	
3	0.889	0.719
7	0.655	0.506
9	0.575	0.469
11	1.028	0.824
14	0.516	0.452
16	0.621	0.623
21	0.496	0.583
25	0.512	0.570
28	0.909	0.844
30	0.770	0.635
	Openness to Experience	
1	0.377	0.427
2	0.725	0.757
5	0.497	0.441
8	0.706	0.486
15	0.700	0.735
20	0.672	0.480
22	0.450	0.383
29	0.749	0.631
37	0.261	0.295
43	0.266	0.407

Table 4
Factor Correlations

Factor	E	A	C	N	O
E					
A	.209				
C	.257	.368			
N	−.370	−.495	−.395		
O	.146	.214	.026	.007	

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

Each model was specified as oblique, with each factor being correlated with each other factor. Given that there are ongoing discussions about whether the factors should be orthogonal or oblique, we also ran an orthogonal model, but we found that the oblique model provided better estimates.[2] The models were fit using maximum likelihood estimation on the covariance structure for only those observations that were complete.

Typically in a CFA analysis, researchers make a choice either to constrain the unstandardized factor pattern coefficients or to constrain the factor variances (MacCallum, 1995). Constraining the unstandardized factor pattern coefficients allows the factor variances to be measured as some function of the constrained variable. However, constraining the factor variances is less restrictive. This method presumes that the variables are independently estimated for different groups and that the same model fits different groups (Thompson, 2004). This typically results in better model fit and allows researchers to interpret the covariances of the factors as factor correlations and to compare the pattern coefficients with each other (Thompson, 2004). Previous studies have constrained factor pattern coefficients (see Lim & Ployhart, 2006); therefore, we constrained the factor variance for each model to 1.00.

Results

Each model had 55 fixed nonzero parameters and used all 760 responses. We began our investigation by estimating a series of reliability coefficients. The responses had good reliability, with Cronbach's alphas of .863, .759, .849, .864, and .778 for E, A, C, N, and O, respectively. These alphas are similar to those found in the literature (e.g., Donnellan et al., 2006; Lim & Ployhart, 2006).

To establish model fit, we estimated several fit statistics, all of which are reviewed in Table 2. We began with the chi-square significance test. Failing to reject the null hypothesis of this test would indicate that the residual covariance estimate equals a matrix that contains only zeros—an indication of perfect model specification. Therefore, rejecting this test would be a sign of bad model fit (Bentler & Bonett, 1980). Unfortunately, this test is sensitive to sample size. Large sample sizes can result in falsely rejecting the null hypothesis (Bentler & Bonett, 1980; Hu & Bentler, 1999; Thompson, 2004). The reverse is also true; smaller sample sizes are much more likely to produce insignificant chi-

[2] Orthogonal model statistics: $\chi^2_{1175} = 5{,}692.346$, $p < .001$, $CFI = .678$, $RMSEA = .071$, $SRMR = .126$.

square values (Bentler & Bonett, 1980). With our very large sample size, the null was rejected in each case. It is noted that the proposed model did have the smallest chi-square value.

We also estimated the CFI, which compares the chi-square between the tested model and the baseline model, or null model, assuming complete independence between the variables. The CFI ranges from 0 to 1, with higher values representing better model fit. Second, we estimated the RMSEA statistic. This statistic estimates how well the population covariances can be reproduced from the model parameters. The RMSEA also has a range from 0 to 1, but here lower values represent better model fit, with an RMSEA of 0 indicating that the estimated model reproduces the population covariances exactly (Hu & Bentler, 1999; Thompson, 2004). Finally, we estimated the SRMR statistic, which is interpreted similarly to the RMSEA, with lower values on the 0 to 1 range representing better model fit.

Bentler (1990) stated that most goodness-of-fit indices are "purely descriptive statistics," with little known about their sam-

Table 5
Exploratory Factor Analysis Varimax Factor Loadings

Item #	Factor 1 (N)	Factor 2 (E)	Factor 3 (O)	Factor 4 (C)	Factor 5 (A)
1	−.050	.156	.431	−.122	.001
2	−.017	−.041	.669	−.014	.104
3	−.627	−.114	.140	−.140	−.138
4	.224	−.296	−.069	.016	.454
5	.015	−.039	.492	−.019	.044
6	.169	.040	−.038	.642	.045
7	−.564	−.004	−.078	−.114	−.054
8	−.024	.098	.573	−.081	−.076
9	−.519	.000	−.029	−.001	−.177
10	.202	.579	−.112	.021	.303
11	−.701	−.138	.108	−.141	−.117
12	.131	−.036	−.110	.593	.013
13	.331	.046	.042	.045	.350
14	−.481	.038	−.013	.068	−.360
15	.036	−.017	.637	.043	.118
16	−.576	−.223	−.048	−.206	−.084
17	.135	.624	.038	.103	−.030
18	.084	.091	.136	.559	.144
19	.018	.070	−.054	.533	−.003
20	−.009	.083	.565	−.033	−.043
21	−.565	−.229	−.080	−.182	−.069
22	−.023	.193	.454	.044	−.012
23	.176	.076	−.045	.584	.192
24	.253	.141	.099	−.015	.386
25	−.482	−.336	.018	−.227	−.092
26	−.021	.599	.147	.148	−.072
27	.305	.569	−.035	.080	.186
28	−.710	−.175	.068	−.156	−.157
29	.034	−.044	.565	.041	.165
30	−.612	.011	.005	−.121	−.201
31	.112	.558	.098	−.056	−.157
32	.143	−.089	.070	.185	.599
33	.135	.116	−.029	.027	.546
34	.140	−.149	.134	.139	.465
35	.126	.176	−.042	.626	.050
36	.119	.401	.186	.186	.121
37	−.049	.439	.318	.136	.026
38	.136	.118	−.010	.674	.090
39	.195	.611	.043	.141	.172
40	.115	.157	.026	.167	.554
41	.010	.065	.035	.495	.101
42	.038	.700	−.004	−.035	−.014
43	.048	.175	.408	.111	.256
44	.019	.130	.173	.070	.481
45	−.007	.680	−.019	.035	−.062
46	.162	−.045	.035	.162	.505
47	.023	.079	−.046	.625	.056
48	.062	.666	.144	.099	.029
49	.128	.425	.064	.106	.383
50	.136	−.008	.133	.569	.117

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

pling distributions. Because of this, few researchers use the same statistics and even when they do, the interpretations of these statistics vary considerably. For example, Lim and Ployhart (2006) used the CFI with a cutoff of greater than .95 and the RMSEA with a cutoff of less than or equal to .08. Thompson (2004) endorsed a .95 cutoff for the CFI and values of .06 and lower for the RMSEA. Donnellan et al. (2006) suggested that their model, which had an RMSEA of .07, was reasonable.

Hu and Bentler (1999) recommended that criteria for a cutoff value should minimize the errors of rejecting a model when it is true (Type I error) while simultaneously minimizing the probability of accepting a model when it is false (Type II error). A researcher should specify a cutoff value accordingly. Through a series of simulations, Hu and Bentler demonstrated that combination rules are optimal because some statistics are most sensitive to models with misspecified factor covariances (such as the SRMR), and others are most sensitive to models with misspecified factor loadings (such as the CFI and RMSEA). We chose to use their combination rules of an SRMR less than .09, in combination with either a CFI greater than or equal to .96 or an RMSEA of less than .06, which they suggested minimizes the sum of both errors.

As can be seen in Table 2, the proposed model has the best goodness-of-fit indices of all the models. This is evidence of construct validity, because the alternate interpretations of the items proposed by the expert panel were not fit by the data as well as the intended interpretation of the items. The alternate model coming closest was the agreeableness model. The SRMR for the proposed model was .083, meeting the cutoff of .09. This suggests that the proposed model is specified well as it pertains to factor covariances. The RMSEA for the proposed model was .068, very slightly over the ideal cutoff of .06. The CFI of .706 fell well below the target value of .96, suggesting that there may be misspecified factor loadings. The proposed model comes very close to meeting the combination rule recommended by Hu and Bentler (1999).

As further evidence of construct validity, we present the individual pattern and structure coefficients in Table 3. Each item loaded significantly ($p < .05$) on its factor. This signifies that each item contributes significantly within its factor. Also important are the correlations between factors (see Table 4), which are consistent with previous studies (e.g., Donnellan et al., 2006).

Because our CFI fell short of the cutoff, we next ran an exploratory factor analysis to investigate potential misspecified factor loadings. We began by extracting the factors using an adjusted principal components analysis and then rotated each factor of the five specified factors using a varimax rotation. Table 5 contains the factor loadings of the rotated solution.

Unfortunately, the literature does not provide an absolute cutoff for the structure coefficients to determine which items are likely to load on each factor if the study were replicated. For example, a cutoff value of |.30| results in Items 10, 13, 14, 25, 27, 37, and 49 loading on multiple factors. If a cutoff of |.35| is used, however, only Items 14 and 49 load on multiple factors. Thompson (2004) suggested that 10 or more structure coefficients around |.40| are needed to define replicable factors when the sample size is greater than 150. Using this cutoff, we found that Items 13 and 24, both of which were supposed to load on A, did not load on any factor. This suggests that A is underrepresented by the item pool. Swapping Items 13 and 24 out with other items might produce a model with stronger fit.

Conclusions

The IPIP project provides an unprecedented and extremely valuable resource for personality researchers, certainly for those working with the five-factor model. The IPIP provides proxy scales for several different five-factor model instruments and marker sets, ranging from the full 300-item set measuring all five domains and 30 narrower facets covered by the NEO-PI–R (Costa & McCrae, 1992), as well as 100-item and 50-item measures of the five domains. Donnellan et al. (2006) presented impressive psychometric data for a 20-item five-factor model measure, with four (IPIP) items per domain. The present study was designed to provide additional reliability and construct validity data for the 50-item IPIP proxy, in a specific ordering and presentation referred to as the M5-50.

The M5-50 has the advantages of being freely and easily accessible via the Internet and being free from the constraints of typical copyrighted personality inventories. Despite this, there have been only a few studies that investigated the reliability and validity of its psychometric properties. We expanded on this research by offering a more diverse sample and a more appropriate CFA methodology. Our results suggest that the M5-50 can be a viable measure of the five-factor model, with room for improvement—particularly on the agreeableness factor.

Although this study has corroborated the construct validity of the M5-50, certainly more work remains to be done. Future studies of the M5-50 should explore more complex models such as those involving cross-loadings. Further, future studies might wish to consider more alternative models. No matter the results, such work would help us better understand how to properly assess personality in a variety of contexts.

References

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology, 90,* 512–527.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

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.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 66,* 93–114.

Cloninger, C. R. (1994). *The Temperament and Character Inventory (TCI): A guide to its development and use.* St. Louis, MO: Washington University, Center for the Psychobiology of Personality.

Conn, S. R., & Rieke, M. L. (1994). *The 16PF fifth edition technical manual.* Champaign, IL: Institute for Personality and Ability Testing.

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.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18,* 192–203.

EQS (Version 6.1) [Computer software]. Encino, CA: Multivariate Software. Available from http://www.mvsoft.com/index.htm

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. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L., Johnson, J., Eber, H., Hogan, R., Ashton, M., Cloninger, R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84–96.

Gough, H. G., & Bradley, P. (1996). *CPI manual* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press.

Hogan, R., & Hogan, J. (1992). *Hogan Personality Inventory manual.* Tulsa, OK: Hogan Assessment Systems.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Kam, C. D., Wilking, J. R., & Zechmeister, E. J. (2007). Beyond the "narrow data base": Another convenience sample for experimental research. *Political Behavior, 29,* 415–440.

Lim, B., & Ployhart, R. E. (2006). Assessing the convergent and discriminant validity of Goldberg's International Personality Item Pool: A multitrait-multimethod examination. *Organizational Research Methods, 9,* 29–54.

MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 16–36). Thousand Oaks, CA: Sage.

McCord, D. M. (2002). M5-50 Questionnaire [Administration and scoring materials]. Retrieved from http://paws.wcu.edu/mccord/m5-50/

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.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515–530.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Received June 11, 2009
Revision received August 7, 2009
Accepted August 10, 2009 ■