The Mini-IPIP Scales: Tiny-Yet-Effective Measures of the Big Five Factors of Personality

M. Brent Donnellan, Frederick L. Oswald, Brendan M. Baird, and Richard E. Lucas
Michigan State University

The Mini-IPIP, a 20-item short form of the 50-item International Personality Item Pool—Five-Factor Model measure (Goldberg, 1999), was developed and validated across five studies. The Mini-IPIP scales, with four items per Big Five trait, had consistent and acceptable internal consistencies across five studies (α at or well above .60), similar coverage of facets as other broad Big Five measures (Study 2), and test-retest correlations that were quite similar to the parent measure across intervals of a few weeks (Study 4) and several months (Study 5). Moreover, the Mini-IPIP scales showed a comparable pattern of convergent, discriminant, and criterion-related validity (Studies 2–5) with other Big Five measures. Collectively, these results indicate that the Mini-IPIP is a psychometrically acceptable and practically useful short measure of the Big Five factors of personality.

Keywords: personality assessment, Big Five, International Personality Item Pool

Investigators often want to measure a wide range of constructs in their research; however, completing a large packet of questionnaires can be a boring or irritating task for participants. This might end up producing transient measurement errors (e.g., Schmidt, Le, & Ilies, 2003) because participants are in a negative mood, or because they respond carelessly due to frustration with the length of the assessment. Moreover, to the extent that it is even mildly unpleasant to participate in research, long questionnaires may increase the likelihood that participants will decide not to complete the study, will drop out of subsequent data collections in longitudinal studies, or will refuse to take part in future studies. Given these kinds of practical concerns, researchers often create shorter forms of longer assessment instruments (Stanton, Sinar, Balzer, & Smith, 2002). However, Smith, McCarthy, and Anderson (2000) cautioned that many well intentioned researchers commit several "sins" in the process of shortening these forms. For instance, although the content is derived from the parent measure, researchers sometimes fail to evaluate whether or not the short form has comparable amounts of reliability and validity and whether psychometric findings for the short form will generalize across independent samples. It is perhaps understandable that these sins are committed given that guidelines concerning the development of short forms are in their infancy (see Marsh, Ellis, Parada, Richards, & Heubeck, 2005; Smith et al., 2000; Stanton et al., 2002); nonetheless, using short forms with poor or poorly understood psychometric properties will likely slow down scientific progress.

In this paper, we describe the process we used to develop and validate a short measure of the Big Five factors of personality (see John & Srivastava, 1999): Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination (or Openness).[1] The Big Five serve as the dominant model of personality structure in trait psychology (e.g., Funder, 2001; but see Block, 1995), and in clinical psychology there is growing interest in using the Big Five to understand Axis I disorders (e.g., Trull & Sher, 1994), Axis II disorders (Costa & Widiger, 2002; Durrett & Trull, 2005; Widiger, 2005; Wiggins, 2003), as well as substance abuse and antisocial behavior (e.g., Ball, 2005; Miller & Lynam, 2001). For example, the premise of the edited book by Costa and Widiger (2002) is that Axis II personality disorders represent constellations of extreme variants of these five broad dimensions of personality. In a related vein, it has often been proposed that basic personality traits like Neuroticism, serve as a diathesis for the development of clinically significant impairments (e.g., Krueger, Caspi, & Moffitt, 2000). The increased interest in integrating models of normal

[1] There is sometimes disagreement over the fifth factor in the Big Five literature. Inventories derived from lexical studies have traditionally labeled the fifth factor Intellect/Imagination, whereas inventories derived from questionnaire items have used the label Openness or Openness to Experience. For the sake of clarity we refer to the fifth factor as Intellect/Imagination when we are generally describing this dimension of personality and when discussing measures that explicitly use that label. We restrict the use of the term Openness to our discussion of inventories that use that label. However, many authors use the terms Openness and Intellect/Imagination more or less interchangeably (e.g., John & Srivastava, 1999).

personality with clinical research on psychopathology (e.g., Durrett & Trull, 2005) represents an important bridge between clinical psychology and personality psychology. Thus, with the increasing interest in personality traits and the Big Five, we anticipate a growing need for short measures of these constructs (e.g., Gosling, Rentfrow, & Swann, 2003).

Why Another Short Big Five Measure?

So why develop another short Big Five measure? This is a reasonable question to ask given that there are several relatively short Big Five measures, with good psychometric characteristics, including the 60-item NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992), the 50-item International Personality Item Pool – Five Factor Model (IPIP-FFM; Goldberg, 1999), the 44-item Big Five Inventory (BFI; John & Srivastava, 1999) and the 40-item Big Five Mini-Markers (Saucier, 1994). These inventories may still be too long, however, particularly in studies where participants will be completing a considerable number of items (e.g., large-scale panel studies) or whenever use of participants' time must be very brief (e.g., experience sampling studies). In fact, these sorts of concerns motivated Gosling et al. (2003) to develop the Ten-Item Personality Inventory (TIPI) measure of the Big Five. The creation of the TIPI is encouraging because it suggests it is possible to measure the Big Five with very few items. However, several considerations motivated us to construct our own short Big Five measure rather than rely on the TIPI.

First and foremost, we suspected that a slightly longer measure of the Big Five would be more practically useful than the TIPI, while still reaping practical benefits over the other short-form personality measures just described. Simply put, we believed that it would be too difficult to obtain adequate internal consistencies and reasonable content or construct breadth with only two items per scale (see Saucier & Goldberg, 2002, for a similar argument). Indeed, three of the five TIPI scales had internal consistency coefficients at or below .50 as reported in Gosling et al. (2003). Moreover, the TIPI might pose serious problems in structural equation modeling contexts because having only two indicators per latent factor can lead to estimation problems and limited modeling flexibility (e.g., Bollen, 1989; Kenny, 1979; Kline, 2004; Little, Lindenberger, & Nesselroade, 1999). This concern over the few number of items on the TIPI also applies to exploratory factor analytic contexts where it is often recommended that each common factor has at least three or four primary indicators (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Floyd & Widaman, 1995; Guadagnoli & Velicer, 1988). Thus, the TIPI faces some serious limitations in latent variable applications.

Second, we specifically wanted our short form to have empirically distinct scales; in other words, we wanted the correlations between the separate scales to be very low. Critics of the Big Five point out nontrivial empirical overlap between Big Five scales that are conceptually orthogonal (see Block, 1995; Funder, 2001). Fortunately, Saucier (2002) demonstrated that this overlap was largely an accidental by-product of scale development techniques that place a premium on maximizing reliability and illustrated that it is possible to create orthogonal and valid measures of the Big Five. The scale intercorrelations for the TIPI were in fact low (e.g., an average correlation of .20 with a maximum of .36 in Study 2 by Gosling et al., 2003); however, this figure might be attenuated

because of the relatively low reliabilities of the scales. We sought to develop a measure with even lower intercorrelations, with more items than the TIPI to increase reliability as well as content and construct breadth, but with much fewer items than Saucier's explicitly orthogonal scales (Saucier, 2002). Thus, we wanted to create short Big Five scales with very low intercorrelations to maximize efficiency in multivariate prediction research. Moreover, empirically distinct scales prevent interpretational problems caused by multicollinearity, which can have "devastating effects on regression statistics to the extent of rendering them useless, even highly misleading" (Pedhazur, 1997, p. 295). Although complex items with content that relates to multiple Big Five constructs can prove to be useful, we wanted our short form to focus on items that were relatively pure indicators of only one of the Big Five constructs.

A third and final consideration is that we were interested in shortening an existing measure so that we could follow many of the guidelines that do exist for developing short forms (Smith et al., 2000; see Marsh et al., 2005, for examples of these procedures). Specifically, we investigated whether the short-form scales have levels of reliability and criterion-related validity that are adequate and similar to the original inventory. Moreover, we wanted to learn about the practical costs to reliability and validity that occur when researchers remove a large number of items from the parent inventory even when following recent "best practice" guidelines.

To summarize, we set out to construct a short inventory of the Big Five with the objective of producing scales that were efficient predictors of meaningful outcomes in psychological research. We elected to shorten the 50-item IPIP-FFM because it is frequently used in personality research and because it is publicly available to researchers on the IPIP website at no cost (http://ipip.ori.org/newQform50b5.htm). We had originally planned to create a 15-item measure but were persuaded by recommendations in Saucier and Goldberg (2002, p. 43–44) that four items serve as "a practical minimum" for scale length. Thus, we opted to create a 20-item measure, with four items per Big Five scale, that we dubbed the Mini-IPIP. As an additional consideration, Saucier and Goldberg (2002) stressed the need for balanced scales that have equal numbers of positively and negatively keyed items. For each Big Five scale, we therefore strived to select 2 items keyed in the negative direction and 2 items keyed in the positive direction, and we were able to meet this goal for all scales except for the Intellect/Imagination scale which comprised 3 negatively keyed items and 1 positively keyed item.

Overview of the Present Studies

We developed and evaluated the 20-item Mini-IPIP inventory across five independent and diverse samples. The 20-item short form was derived in Studies 1 and 2. Study 1 details the development of the Mini-IPIP using a very large sample of students from multiple colleges and universities ($N = 2,663$). We used Study 2 to fine-tune the item selection for the Mini-IPIP and to examine how well the Mini-IPIP relates to Big Five facets, the IPIP-FFM, and the TIPI using a sample (undergraduates in psychology courses) and sample size ($N = 329$) that is similar to those typically used in psychological research. That is, we wanted to know how the Mini-IPIP related to facets of the Big Five as well

as other broad measures of the Big Five. Study 3 was conducted to replicate findings for the psychometric properties of the Mini-IPIP on an independent sample to ensure that the desirable properties found in the items selected for the measure were not a result of capitalizing on chance. Study 3 also provided data on how well the Mini-IPIP related to an alternative Big Five measure and to several criterion measures. Studies 4 and 5 were based on datasets not originally designed to validate the short scales. Nonetheless, these datasets provided short-term (Study 4) and longer-term (Study 5) test–retest reliability information for the Mini-IPIP as well as more criterion-related validity evidence. Moreover, Study 5 provided data on how well the Mini-IPIP correlated with informant-reports of personality. As a package, these five studies reflect the simple yet powerful technique of replication to demonstrate that the 20-item Mini-IPIP is a reasonable replacement for the parent inventory in situations where the length of the personality assessment is a major concern.

Study 1: Development of the 20-item Mini-IPIP Inventory

Sample and Measures

Complete IPIP-FFM data were available from 2,663 freshman undergraduate students across 10 colleges and universities involved in a large-scale project designed to study how well individual difference measures worked as predictors of academic performance. The substantive work from that project is detailed elsewhere (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). Regarding the demographic breakdown, 97% of the sample was either 18 or 19 years of age, 64% of the sample was female, 96% were U.S. citizens, and 94% indicated that English was their first language. The racial composition for this sample was 55% Caucasian, 25% African American, 6% Hispanic, 7% Asian, and 7% other ethnicities. Participants completed the 50-item IPIP-FFM measure (Goldberg, 1999) by indicating, on a 5-point scale, how well each statement described them. Descriptive statistics for each of the 10-item subscales are displayed in Table 1.

Results and Discussion

Method for Creating the Mini-IPIP Scales

Although there are many approaches to selecting items for short forms, we elected to use a simple approach that can be readily

implemented with available statistical packages. We conducted a principal axis exploratory factor analysis (EFA) with a varimax rotation on the 50-item IPIP-FFM dataset. We forced a five-factor solution and exported the factor-loading matrix to a spreadsheet program. As expected, all of the items loaded appropriately on their corresponding factors; however some loaded more highly than others, and some had large cross-loadings (i.e., nontrivial loadings on multiple factors). In the next step, for each IPIP-FFM item, we calculated the difference between its loading on its primary factor and the average of the absolute "off" factor loadings. We labeled this the "discrimination score" for each item. For example, the first IPIP-FFM item, "I am the life of the party," primarily loads on the Extraversion factor, so we subtracted the average of the absolute value of its loadings on the other factors (the average of the absolute value of the loadings on the Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination factors = .19) from its primary loading on the Extraversion factor (loading = .64 to yield a discrimination score of .45).

After calculating discrimination scores for each item, we selected the two positively and two negatively keyed items from each scale that had the largest discrimination scores. (Regarding the latter, recall that one of our goals in creating the Mini-IPIP was to create scales that were empirically distinct from one another.) It is important to note that we were limited in item selection for the Neuroticism and Intellect/Imagination scales because the IPIP-FFM contains only two negatively keyed Neuroticism items and three negatively keyed Intellect/Imagination items.

We then subjected our initial 20-item pool to a separate EFA on the sample in Study 2 (to be described subsequently). Our goal was to see if we could recover a five-factor structure with reasonable factor loadings for each of the items. We were able to get a clear five-factor solution, but we had to make modifications to two of our original scales because some items did not have particularly strong factor loadings. First, for Conscientiousness scale, the initially selected item "I follow a schedule" (Original IPIP-FFM item #43; discrimination score = .38) had a relatively low loading in the new sample (.33). Thus, we replaced it with the item, "I like order" (Original IPIP-FFM item #33; discrimination score = .37) because this item had the next highest discrimination score of the positively keyed items.

Table 1

Descriptive Statistics for the IPIP-FFM and Mini-IPIP (Studies 1 and 2)

	Convergent Validity	IPIP-FFM (50-items)			Mini-IPIP (20-items)		
		α	Mean	SD	α	Mean	SD
Study 1 (N = 2,663)							
Extraversion	.93	.87	3.36	.77	.77	3.28	.90
Agreeableness	.89	.80	4.00	.57	.70	4.01	.69
Conscientiousness	.90	.80	3.57	.61	.69	3.42	.78
Neuroticism	.92	.85	2.72	.73	.68	2.54	.80
Intellect/Imagination	.85	.80	3.63	.58	.65	3.70	.73
Study 2 (N = 329)							
Extraversion	.94	.91	3.50	.81	.82	3.45	.90
Agreeableness	.91	.80	4.10	.54	.75	4.15	.64
Conscientiousness	.90	.81	3.49	.63	.75	3.40	.86
Neuroticism	.93	.89	2.84	.83	.70	2.62	.83
Intellect/Imagination	.83	.78	3.68	.57	.70	3.74	.76

Second, for the Intellect/Imagination scale, we had to replace the items "I have excellent ideas" (Original IPIP-FFM item #25; discrimination score = .40) and "I use difficult words" (Original IPIP-FFM item #40; discrimination score = .42) because neither loaded on the primary factor in the new sample above .30. The Intellect/Imagination scale was the only instance where we deviated from the procedure of selecting items solely on the basis of the discrimination scores. We deviated from this procedure because we wanted to select items with content that seemed distinct from those items that appeared to measure general intelligence and even narcissism. We first selected the item "I have a vivid imagination" (Original IPIP-FFM item #15; discrimination score = .38) because it was the positively keyed item with the fourth highest discrimination score. The item "I am full of ideas" (Original IPIP-FFM item #50; discrimination score = .40) had a higher discrimination index, but we eliminated it in favor of the former item because the former item seemed like a purer measure of imagination. We then selected the last negatively keyed Intellect/Imagination item, "I do not have a good imagination" (Original IPIP-FFM item #30; discrimination score = .26) because the remaining positively keyed items that had decent discrimination scores and primary factor loadings also seemed to tap general intelligence and/or narcissism (i.e., "I have a rich vocabulary" and "I am quick to understand things"). Appendix A lists the 20 items selected for the Mini-IPIP scales, where the item numbers refer to the order of the items in the version of the 50-item IPIP-FFM inventory available at the IPIP website.

Basic Psychometric Properties of the Reduced Scales

Table 1 presents means, standard deviations, and coefficient alphas for the Mini-IPIP scales. As seen in Table 1, the Mini-IPIP scales had acceptable reliability especially in light of their reduced length. These coefficients ranged from .65 for Intellect/Imagination to .77 for Extraversion. Table 1 also shows the convergent correlations between the Mini-IPIP scales and the 10-item "parent" IPIP-FFM scales. These were high, ranging from .85 for Intellect/Imagination to .93 for Extraversion. These correlations are inflated by the fact that each pair of scales contains four identical items, but from the practical standpoint of comparing the shorter with the longer scale, they are still informative. That said, we computed the associations between the Mini-IPIP scales and the 6-items not included in the scale, and these correlations were also high (Extraversion = .78, Agreeableness = .67, Conscientiousness = .67, Neuroticism = .76, and Intellect/Imagination = .56).

Regarding discriminant validity, the scale intercorrelations were successfully reduced for the Mini-IPIP scales as compared to the IPIP-FFM scales. Specifically, the average absolute scale intercorrelation for the Mini-IPIP was $r = .13$ ($SD = .08$, Range = .02–.24) compared to an average absolute scale intercorrelation for the full IPIP-FFM of $r = .20$ ($SD = .08$, Range = .07–.35). We repeated these analyses correcting the intercorrelations for attenuation due to measurement error, and the results were similar (Mini-IPIP: average $r = .18$, $SD = .11$, Range = .02–.34; compared to the IPIP-FFM: average $r = .25$, $SD = .10$, Range = .08–.41).

Study 2: An Examination of the Content Validity of the IPIP-FFM and Mini-IPIP Scales

Sample

This sample consists of 329 undergraduate students enrolled in psychology courses at a large public research university in Michigan who participated in exchange for course credit or extra credit during the Fall Semester of 2005. Data were collected over the Internet using a web-based interface. Participants were primarily female (73.3%) and either first-year or second-year college students (33.3% and 35.5% respectively). Four additional participants' data were discarded because they failed to answer "Yes" to the final statement: "I answered all of these questions honestly."

IPIP-FFM and Mini-IPIP Measures

IPIP-FFM and Mini-IPIP. The bottom half of Table 1 presents descriptive statistics and convergent scale intercorrelations for the 50-item IPIP-FFM and the 20-item Mini-IPIP for Study 2. In general, the values of these coefficients are close to those in Study 1. The average absolute scale intercorrelation for the Mini-IPIP scales was $r = .14$ ($SD = .08$; maximum = .30 for Agreeableness and Intellect/Imagination) compared to $r = .21$ for the IPIP-FFM scales ($SD = .11$; maximum = .36 for Agreeableness and Intellect/Imagination). We repeated these analyses correcting the intercorrelations for attenuation due to measurement error, and the results were similar (Mini-IPIP: average $r = .19$, $SD = .11$; IPIP-FFM: average $r = .25$, $SD = .13$).

Comparison Measures

The IPIP-NEO. Participants completed the 120-item IPIP measure developed by Johnson (2000) to measure the same facets of the Big Five that are assessed by the NEO-PI (Costa & McCrae, 1992). The IPIP-NEO includes items that measure 30 facets of the Big Five– 4 items per facet and 6 facets per factor— on a 5-point scale. Internal consistencies for the facet scales are reported in Table 2. Scores for the Big Five were calculated by taking the average of all of the respective facet-level items for each of the Big Five: Extraversion: $M = 3.53$, $SD = .51$, $\alpha = .87$; Agreeableness: $M = 3.75$, $SD = .46$, $\alpha = .84$; Conscientiousness: $M = 3.62$, $SD = .54$, $\alpha = .89$; Neuroticism: $M = 2.77$, $SD = .61$, $\alpha = .90$; Openness: $M = 3.40$, $SD = .44$, $\alpha = .78$. Perhaps with the exception of Openness, these Big Five scores demonstrated reasonable convergent validity with the IPIP-FFM ($r = .75$, .61, .75, .79, .55, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness with Intellect/Imagination, respectively) and the Mini-IPIP ($r = .70$, .52, .63, .73, .55, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness with Intellect/Imagination, respectively).

Ten-Item Personality Inventory (TIPI). Participants completed the TIPI using a 7-point scale, with the following results: Extraversion: $M = 4.98$, $SD = 1.45$, $\alpha = .70$ (r for the 2 items = .56); Agreeableness: $M = 5.38$, $SD = 1.10$, $\alpha = .32$ (r for the 2 items = .22); Conscientiousness: $M = 5.55$, $SD = 1.17$, $\alpha = .43$

Table 2
Big Five Facet Coverage by Measures of the Big Five Factors (Study 2)

	IPIP-FFM Scale	Mini-IPIP Scale	TIPI Scale
Extraversion			
Friendliness (.73)	.67	.60	.60
Gregariousness (.70)	.69	.70	.57
Assertiveness (.85)	.52	.45	.46
Activity Level (.65)	.24	.23	.20
Excitement Seeking (.63)	.45	.44	.40
Cheerfulness (.78)	.44	.39	.38
Agreeableness			
Trust (.85)	.32	.22	.38
Morality (.75)	.48	.41	.33
Altruism (.66)	.68	.64	.36
Cooperation (.72)	.34	.25	.39
Modesty (.70)	.06	.05	.04
Sympathy (.68)	.52	.52	.21
Conscientiousness			
Self-Efficacy (.79)	.45	.28	.38
Orderliness (.82)	.72	.74	.57
Dutifulness (.56)	.42	.33	.43
Achievement Striving (.79)	.52	.36	.54
Self-Discipline (.69)	.62	.50	.54
Cautiousness (.85)	.37	.31	.40
Neuroticism			
Anxiety (.77)	.72	.61	.57
Anger (.83)	.60	.53	.54
Depression (.84)	.65	.65	.56
Self-Consciousness (.72)	.28	.28	.22
Immoderation (.66)	.31	.28	.26
Vulnerability (.74)	.68	.64	.63
Openness			
Imagination (.68)	.45	.47	.32
Artistic Interest (.66)	.40	.43	.35
Emotionality (.55)	.16	.15	.19
Adventurousness (.60)	.26	.26	.40
Intellect (.67)	.59	.53	.32
Liberalism (.66)	.05	.10	.13

Note. $N = 329$. Numbers inside parentheses are the alpha reliability estimates for the 4-item facet scales of the 120-item IPIP-NEO. TIPI = Ten Item Personality Inventory (Gosling, Rentfrow, & Swann, 2003)

(r for the 2 items = .29); Neuroticism: $M = 3.17$, $SD = 1.51$, $\alpha = .69$ (r for the 2 items = .54); Openness: $M = 5.51$, $SD = 1.07$, $\alpha = .41$ (r for the 2 items = .27). With the possible exception of Agreeableness and Openness, the TIPI demonstrated reasonable convergent validity with the IPIP-NEO ($r = .65, .46, .69, .67, .48$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness, respectively), the IPIP-FFM ($r = .79, .41, .70, .74, .52$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness with Intellect/Imagination, respectively), and the Mini-IPIP ($r = .75, .33, .63, .73, .46$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness with Intellect/Imagination, respectively).

Results and Discussion

Content Validity of the IPIP-FFM, Mini-IPIP, and TIPI Scales

Table 2 displays the correlations between the Big Five scales assessed by the IPIP-FFM, Mini-IPIP, and TIPI and the corre-

sponding facets of that factor as assessed by the IPIP-NEO. For instance, Table 2 shows how these three measures of Extraversion correlate with the six Extraversion facets of Friendliness, Gregariousness, Assertiveness, Activity Level, Excitement Seeking, and Cheerfulness. Generally speaking, the pattern of correlations was very similar across all three measures. More specific to our interests, the content coverage of the Mini-IPIP was very close to coverage of the IPIP-FFM. For instance, across the 30 facets there were only 2 cases where the difference in magnitude between the IPIP-FFM and for the Mini-IPIP was greater than |.12|. This occurred for the facets of Achievement Striving (difference = .16), and Self-Efficacy (difference = .17).

Study 3: Replication of the Psychometric Properties and Evidence of the Criterion-Related Validity of the Mini-IPIP Scales

Sample

This sample consists of 300 undergraduate students enrolled in psychology courses at a large public research university in Michigan who participated in exchange for course credit or extra credit during the Spring Semester of 2005. Data were collected over the Internet using a web-based interface. Participants were primarily female (78.7%) and identified themselves as European American (63.3%), Asian American (5.0%), African American (2.3%), Latino/a (2.0%), or self-reported they were a member of some "Other" ethnic group (27.3%). Data from three additional participants were discarded because they failed to answer "Yes" to the final statement: "I answered all of these questions honestly".

IPIP-FFM and Mini-IPIP. Participants completed the 50-item IPIP-FFM measure, with the following results: Extraversion: $M = 3.48$, $SD = .73$, $\alpha = .90$; Agreeableness: $M = 4.02$, $SD = .52$, $\alpha = .83$; Conscientiousness: $M = 3.53$, $SD = .60$, $\alpha = .83$; Neuroticism: $M = 2.87$, $SD = .76$, $\alpha = .89$; Intellect/Imagination: $M = 3.55$, $SD = .52$, $\alpha = .79$. Statistics for the 20-item Mini-IPIP were then calculated from the relevant subset of items within this measure: $M = 3.46$, $SD = .82$, $\alpha = .82$; Agreeableness: $M = 4.06$, $SD = .61$, $\alpha = .77$; Conscientiousness: $M = 3.48$, $SD = .76$, $\alpha = .74$; Neuroticism: $M = 2.70$, $SD = .61$, $\alpha = .78$; Intellect/Imagination: $M = 3.64$, $SD = .67$, $\alpha = .70$. Once again, there was good convergence between the IPIP-FFM and the Mini-IPIP for each of the Big Five ($r = .95, .88, .92, .93, .85$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively).

Convergent Validity and Criterion-Related Validity Measures

The Big Five Inventory (BFI). Participants completed this 44-item measure which also assesses the Big Five traits (John & Srivastava, 1999) using a 5-point scale: Extraversion (8 items): $M = 3.43$, $SD = .72$, $\alpha = .86$; Agreeableness (9 items): $M = 3.82$, $SD = .56$, $\alpha = .79$; Conscientiousness (9 items): $M = 3.63$, $SD = .60$, $\alpha = .83$; Neuroticism (8 items): $M = 2.93$, $SD = .73$, $\alpha = .84$; Openness (10 items): $M = 3.50$, $SD = .57$, $\alpha = .79$. The BFI demonstrated good convergent validity with the IPIP-FFM ($r =$

.84, .64, .73, .86, .74 for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness with Intellect/Imagination, respectively) and the Mini-IPIP ($r = .81, .49, .66, .80, .68$ for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness with Intellect/Imagination, respectively).

Self-esteem. Participants completed the 10-item Rosenberg Self-Esteem Scale using a 5-point scale (Rosenberg, 1965; $M = 3.84$, $SD = .66$, $\alpha = .88$). Robins, Tracy, Trzesniewski, Potter, and Gosling (2001) presented evidence that the Big Five are linked with global self-esteem. Specifically, they reported modest to strong correlations between self-esteem and Neuroticism and between self-esteem and Extraversion in an extremely large Internet sample (see also Schmitt & Allik, 2005). They also reported small to modest positive correlations between self-esteem and the three other Big Five factors. Thus, we expected a similar pattern of associations with the Big Five measures in Study 3. One participant did not complete the self-esteem items, so the sample size was 299 for this scale.

Behavioral inhibition/behavioral approach (BIS/BAS). Respondents completed the original Carver and White (1994) measures using a 5-point scale. The Behavioral Inhibition Scale (BIS) measures individual differences in the sensitivity to the behavioral avoidance or inhibition system (7 items: $M = 3.67$, $SD = .62$, $\alpha = .79$; sample item: "Criticism or scolding hurts me quite a bit"), whereas the Behavioral Approach scale (BAS) measures individual differences in the sensitivity to the behavioral approach system (13 items: $M = 3.67$, $SD = .50$, $\alpha = .85$; sample item: "It would excite me to win a contest"). The behavioral inhibition system is believed to underlie Neuroticism primarily (Carver & White, 1994; Watson, Wiese, Vaidya, & Tellegen, 1999), whereas the behavioral approach system is believed to underlie Extraversion (Lucas, Diener, Grob, Suh, & Shao, 2000; Watson et al., 1999). Thus, the Neuroticism and Extraversion scales of the Big Five measures should relate to the BIS and BAS scales, respectively, along these same lines.

Results and Discussion

Confirmatory Factor Analysis of the 4-Item IPIP Scales

We conducted a confirmatory factor analysis (CFA) on the structure of the Mini-IPIP using the program AMOS 5.0 (Arbuckle, 2003; variance-covariance matrix available from the first author). We restricted these analyses to those 296 participants with complete item level data. A recent CFA of the IPIP-FFM did not show good model fit at the item level (Lim & Ployhart, 2006), and indeed, the fit of the CFA model for IPIP-FFM for this sample was certainly not ideal ($\chi^2 = 2,822.68$, $df = 1,165$, $p < .05$; $\chi^2/df = 2.42$; CFI = .74; RMSEA = .07, p close fit $< .05$; parameter estimates for this model are available from the first author). Past published CFA models of Big Five inventories have had to estimate secondary loadings to obtain even remotely satisfactory indices of overall model fit (e.g., Church & Burke, 1994; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). Thus, it might not be possible to obtain reasonable fit from a CFA perspective on many or even most Big Five inventories. Much of the model misfit in omnibus inventories arises from the fact that, "most items and tests tend to have substantial relations with at least two factors rather than with only one" (Goldberg, 1993. p. 186).

Keeping these caveats in mind, we tested the structure of the Mini-IPIP using a CFA model. The fit of the CFA model was rejected by the chi-square test of exact fit (which is sensitive to even trivial misspecifications with increasing sample sizes); however, the fit of the model was within reason based on the RMSEA value ($\chi^2 = 359.30$, $df = 160$, $p < .05$; $\chi^2/df = 2.25$; CFI = 0.88; RMSEA = 0.07, p close fit $< .05$). Relevant coefficients from this CFA model are reported in Table 3. The only potentially low loading was for the Neuroticism item "I seldom feel blue" (standardized loading = .39). We also examined the model modification indices and found suggestions that modeling several cross-loadings for specific items would improve model fit. However, given that many parameter estimates obtained through model modification fail to replicate (e.g., MacCallum, 1986) and the fact that we found reasonable fit without the modifications, we did not choose to model secondary factor loadings post hoc. Our results suggest that the Mini-IPIP is at least as useful as other Big Five measures for CFA models.

Criterion-Related Validity for the Big Five Measures

Table 4 displays the correlations between the Big Five measures and the three criterion measures used in Study 3. Patterns of criterion-related validities were similar across all Big Five measures, with the exception of the Agreeableness-BIS relationship being positive for the IPIP-FFM and Mini-IPIP and near-zero for the BFI. Differences between the IPIP-FFM and the Mini-IPIP scales were not substantial from a practical standpoint. As a final comparison between measures, we conducted a series of regression analyses to examine the predictive validity of each of the Big Five measures as a set. For each analysis, we regressed each of the three criteria on a given set of Big Five scales and recorded the multiple R values (see the bottom of Table 4). In general, the multiple R values were very similar across the three Big Five measures.

Study 4: Short-Term Retest Reliability and Additional Evidence for Criterion-Related Validity of the Mini-IPIP Related to Psychopathology

Sample

Test-retest data were available from 216 undergraduate students enrolled in psychology courses at a large research university in Michigan who participated in exchange for course credit or extra credit during the Spring Semester of 2005. Data were collected over the Internet using a web-based interface. The interval between the Time 1 and Time 2 assessments was approximately three weeks. All psychopathology-related criterion-related measures were administered at Time 2. Information on gender and racial/ethnic group membership was not collected, but the participant pool was the same as Study 2 and Study 3, and therefore demographic characteristics are likely to be quite similar. These data were initially collected as part of a different study investigating response biases in daily self reports.

IPIP-FFM and Mini-IPIP. Participants completed the 50-item IPIP-FFM measure at Time 1, with the following descriptive statistics: Extraversion: $M = 3.41$, $SD = .73$, $\alpha = .88$; Agreeableness: $M = 4.11$, $SD = .52$, $\alpha = .80$; Conscientiousness: $M = 3.51$,

Table 3
Confirmatory Factor Analysis of the Mini-IPIP (Study 3)

Mini-IPIP Item Number	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Intellect/ Imagination
			Standardized Loading		
1	.68				
6	.76				
11	.74				
16	.75				
2		.76			
7		.56			
12		.75			
17		.72			
3			.65		
8			.67		
13			.59		
18			.67		
4				.80	
9				.58	
14				.80	
19				.39	
5					.68
10					.50
15					.52
20					.72

	1.	2.	3.	4.	5.
			Correlations between Latent Variables		
1. Extraversion	1.00				
2. Agreeableness	.30	1.00			
3. Neuroticism	−.21	−.18	1.00		
4. Conscientiousness	.09	.12	−.22	1.00	
5. Intellect/Imagination	.35	.40	−.18	.02	1.00

Note. $N = 296$. Model fit results: $\chi^2 = 359.30$, $df = 160$, $p < .05$; $\chi^2/df = 2.25$; CFI = 0.88; RMSEA = 0.07, p close fit $< .05$.

$SD = .58$, $\alpha = .77$; Neuroticism: $M = 2.85$, $SD = .80$, $\alpha = .89$; Intellect/Imagination: $M = 3.64$, $SD = .56$, $\alpha = .79$. The 20-item Mini-IPIP was then calculated from the relevant subset of items within this measure: $M = 3.39$, $SD = .87$, $\alpha = .81$; Agreeableness: $M = 4.16$, $SD = .61$, $\alpha = .69$; Conscientiousness: $M = 3.42$, $SD = .73$, $\alpha = .60$; Neuroticism: $M = 2.67$, $SD = .85$, $\alpha = .76$; Intellect/Imagination: $M = 3.72$, $SD = .75$, $\alpha = .70$. There was good convergence between the IPIP-FFM and the Mini-IPIP for each of the Big Five at Time 1 ($r = .94$, .88, .89, .94, .87, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively). Participants completed the 50-item IPIP-FFM measure at Time 2: Extraversion: $M = 3.46$, $SD = .77$, $\alpha = .90$; Agreeableness: $M = 4.11$, $SD = .54$, $\alpha = .81$; Conscientiousness: $M = 3.58$, $SD = .59$, $\alpha = .77$; Neuroticism: $M = 2.97$, $SD = .83$, $\alpha = .90$; Intellect/Imagination: $M = 3.75$, $SD = .59$, $\alpha = .81$. As before, the 20-item Mini-IPIP was then calculated from the relevant subset of items within this measure: Extraversion: $M = 3.47$, $SD = .90$, $\alpha = .83$; Agreeableness: $M = 4.20$, $SD = .63$, $\alpha = .72$; Conscientiousness: $M = 3.52$, $SD = .76$, $\alpha = .63$; Neuroticism: $M = 2.78$, $SD = .85$, $\alpha = .75$; Intellect/Imagination: $M = 3.77$, $SD = .75$, $\alpha = .67$. There was good convergence between the IPIP-FFM and the Mini-IPIP for each of the Big Five at Time 2 ($r = .94$, .89, .88, .94, .87, for Extraversion,

Agreeableness, Conscientiousness, Neuroticism, and Intellect/ Imagination, respectively).

Measures of Psychopathology-Relevant Criteria at Time 2

Anxiety. Participants completed the 20-item Spielberger Trait Anxiety Scale (Spielberger, 1983) using a 4-point scale ($M = 2.08$, $SD = .51$, $\alpha = .92$).

Depression. Participants completed the recently developed 10-item short form of the Center for Epidemiological Studies Depression scale using a 4-point scale (Cole, Rabin, Smith, & Kaufman, 2004; M = 1.78, SD = .48, $\alpha = .80$).

Aggression/hostility. Participants completed the 29-item Buss-Perry Aggression Questionnaire using a 5-point scale (Buss & Perry, 1992; M = 2.21, SD = .57, $\alpha = .91$). We computed a total score by taking the average of all of the items.

Psychological entitlement. Participants completed the 9-item inventory developed by Campbell, Bonacci, Shelton, Exline, and Bushman (2004) using a 5-point scale ($M = 2.45$, $SD = .75$, $\alpha = .87$). This measure assesses the degree to which individuals feel more deserving than others, an individual difference that in past

Table 4

Criterion-Related Validity for Big-Five Measures for Predicting Self-Esteem and Behavioral Approach/Avoidance (Study 3)

	Criteria		
Big Five Scale	Self-Esteem	BIS	BAS
Extraversion			
BFI	.36	−.14	.50
IPIP-FFM	.38	−.12	.52
Mini-IPIP	.35	−.13	.50
Agreeableness			
BFI	.37	.06	.11
IPIP-FFM	.36	.21	.28
Mini-IPIP	.24	.26	.23
Conscientiousness			
BFI	.38	−.02	.13
IPIP-FFM	.30	−.03	.14
Mini-IPIP	.26	−.07	.04
Neuroticism			
BFI	−.55	.57	−.08
IPIP-FFM	−.59	.51	−.02
Mini-IPIP	−.60	.42	−.03
Imagination/Intellect			
BFI	.23	−.11	.35
IPIP-FFM	.37	−.12	.43
Mini-IPIP	.28	−.14	.29
R			
BFI	.67	.64	.55
IPIP-FFM	.72	.64	.60
Mini-IPIP	.68	.57	.54

Note. *N* = 299 to 300. BFI = Big Five Inventory (John & Srivastava, 1999), BIS = Behavioral Inhibition Scale (Carver & White, 1994), BAS = Behavioral Approach System (Carver & White, 1994).

research has been associated with unpleasant interpersonal behaviors, such as taking candy that was intended to be given to young children (Campbell et al., 2004). As to be expected, this scale was positively correlated with the Aggression/Hostility scale in these data (*r* = .32).

Results and Discussion

Short-Term Retest Correlations

The short-term retest correlations for the IPIP-FFM scales were high: *r* = .89, .72, .79, .87, and .83, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination,respectively. The retest correlations for Mini-IPIP scales were also high: *r* = .87, .62, .75, .80, and .77, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively.

Criterion-Related Validity

Table 5 displays the validity coefficients for the Big Five assessed at Time 1 predicting scores on criterion measures assessed at Time 2, approximately three weeks later. In general, the coefficients were similar for the IPIP-FFM and Mini-IPIP scales, further indicating that the Mini-IPIP displays a similar pattern of criterion-related validity. Lastly, the total amount of variance explained by the Big Five was quite similar for the IPIP-FFM scales and the Mini-IPIP scales, as shown by the multiple *R*s.

Study 5: The Mini-IPIP Scales' Longer-Term Stability, Associations With Informant Reports, and Further Evidence of Criterion-Related Validity Related to Affect and Life Satisfaction

Method

Participants

Data for this study come from Study 3 in Baird, Le, and Lucas (2006). The present analyses are unrelated to the aims of Baird et al., and the most substantial overlap is that retest correlations for the IPIP are also contained in that report. Nonetheless, this dataset is particularly important for evaluating the Mini-IPIP because it provides longitudinal data collected over a six- to nine-month period in addition to an informant report of personality. Retest data were available from 148 undergraduates (78% women) who were recruited for a study of personality and affect that took place over the nine months of the school year at a large public university in Michigan. Data analyzed here are based on questionnaire assessments that were completed at both time points.

Each participant was asked to provide the names of friends or family members who could serve as informants. Informants rated participants' personality using the IPIP-FFM items and were given the choice of responding via web site or a paper-and-pencil questionnaire. Informants who completed the questionnaire were entered into a drawing for $50. There were an average of 3.08 reports per participant (*SD* = 1.06), and we obtained at least one informant report for 133 participants (90% of the 148). For individuals with more than one informant report (91% of the 133), we averaged item responses across informants before creating aggregate scales.

Personality Measures

IPIP-FFM and Mini-IPIP. Participants completed the 50-item IPIP-FFM measure at Time 1, with the following descriptive statistics: Extraversion: *M* = 3.58, *SD* = .83, α = .91; Agreeableness: *M* = 4.25, *SD* = .50, α = .81; Conscientiousness: *M* = 3.61,

Table 5

Criterion-Related Validity for Big-Five Measures Predicting Psychopathology-Relevant Outcomes (Study 4)

	Criteria at Time 2 (≈3 Weeks Later)			
Big Five Scale at Time 1	Anxiety	Depression	Hostility/ aggression	Psychological entitlement
Extraversion				
IPIP-FFM	−.30	−.24	−.07	.22
Mini-IPIP	−.22	−.20	−.06	.19
Agreeableness				
IPIP-FFM	−.17	−.16	−.33	−.16
Mini-IPIP	−.18	−.15	−.26	−.12
Conscientiousness				
IPIP-FFM	−.20	−.22	−.22	.02
Mini-IPIP	−.19	−.16	−.21	.03
Neuroticism				
IPIP-FFM	.80	.54	.48	.03
Mini-IPIP	.77	.53	.49	.01
Imagination/Intellect				
IPIP-FFM	−.14	−.15	.01	.07
Mini-IPIP	−.07	−.12	.03	.06
R				
IPIP	.82	.58	.59	.34
Mini-IPIP	.79	.56	.56	.28

Note. *N* = 216.

$SD = .64$, $\alpha = .82$; Neuroticism: $M = 2.99$, $SD = .87$, $\alpha = .91$; Intellect/Imagination: $M = 3.86$, $SD = .52$, $\alpha = .79$. Statistics for the 20-item Mini-IPIP were then calculated from the relevant subset of items within this measure: Extraversion: $M = 3.55$, $SD = .94$, $\alpha = .83$; Agreeableness: $M = 4.32$, $SD = .57$, $\alpha = .72$; Conscientiousness: $M = 3.49$, $SD = .84$, $\alpha = .73$; Neuroticism: $M = 2.79$, $SD = .98$, $\alpha = .83$; Intellect/Imagination: $M = 3.98$, $SD = .67$, $\alpha = .69$. There was good convergence between the IPIP-FFM and the Mini-IPIP for each of the Big Five at Time 1 ($r = .94, .87, .91, .96, .86$, for Extraversion, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively). Participants completed the 50-item IPIP-FFM measure at Time 2, with the following descriptive statistics: Extraversion: $M = 3.59$, $SD = .83$, $\alpha = .92$; Agreeableness: $M = 4.25$, $SD = .54$, $\alpha = .87$; Conscientiousness: $M = 3.61$, $SD = .67$, $\alpha = .85$; Neuroticism: $M = 2.93$, $SD = .83$, $\alpha = .90$; Intellect/Imagination: $M = 3.87$, $SD = .53$, $\alpha = .80$. Again, statistics for the 20-item Mini-IPIP were then calculated from the relevant subset of items within this measure: $M = 3.59$, $SD = .94$, $\alpha = .87$; Agreeableness: $M = 4.30$, $SD = .56$, $\alpha = .77$; Conscientiousness: $M = 3.52$, $SD = .85$, $\alpha = .76$; Neuroticism: $M = 2.74$, $SD = .90$, $\alpha = .78$; Intellect/Imagination: $M = 3.94$, $SD = .67$, $\alpha = .70$. There was good convergence between the IPIP-FFM and the Mini-IPIP for each of the Big Five at Time 2 ($r = .95, .91, .92, .96, .86$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively).

Informant reports. Informants rated participants on the 50-item IPIP-FFM scales, with the following descriptive statistics: Extraversion: $M = 3.70$, $SD = .64$, $\alpha = .92$; Agreeableness: $M = 4.15$, $SD = .49$, $\alpha = .90$; Conscientiousness: $M = 3.62$, $SD = .60$, $\alpha = .91$; Neuroticism: $M = 2.85$, $SD = .61$, $\alpha = .90$; Intellect/Imagination: $M = 3.91$, $SD = .43$, $\alpha = .86$. Note that reliabilities are relatively high because the "items" are actually aggregated reports across multiple informants.

Criterion Variables at Time 2

Positive affect. Participants completed the 8-item scale from the Intensity and Time Affect Survey (ITAS; Diener, Smith, & Fujita, 1995) using a 5-point scale ($M = 3.85$, $SD = .65$, $\alpha = .87$).

Negative affect. Participants completed the 16-item scale from the ITAS ($M = 2.21$, $SD = .70$, $\alpha = .93$).

Life satisfaction. Participants completed the 5-item Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffen, 1985) using a 5-point scale ($M = 3.61$, $SD = .98$, $\alpha = .86$).

Results and Discussion

Longer-Term Retest Correlations

The longer-term retest correlations for the IPIP-FFM scales were very high: $r = .88, .79, .81, .82$, and $.82$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively. The retest correlations for Mini-IPIP scales were also high: $r = .86, .68, .77, .82, .75$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively. Thus, the longer-term retest reliability of Mini-IPIP scales resembles the coefficients of the original IPIP-FFM scales.

Associations With Informant Reports

The IPIP-FFM scales at Time 1 were associated with informant reports of the Big Five: $r = .52, .31, .51, .34$, and $.35$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively. Likewise, the Mini-IPIP scales at Time 1 were associated with informant reports of the Big Five: $r = .53, .30, .47, .37$, and $.26$, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Intellect/Imagination, respectively. Thus, the Mini-IPIP scales had similar properties as the IPIP-FFM scales in terms of the associations between self reports and informant reports of personality. Note that these correlations do not approach unity, nor were they expected to, given that self-other correlations for personality traits are often in the range of .30 to .60 (e.g., Funder, 1999).

Criterion-Related Validity

Table 6 displays the correlations between the Big Five assessed at Time 1 and criterion measures assessed at Time 2. As in the previous studies, the coefficients were similar for the IPIP-FFM and Mini-IPIP scales, further indicating that the Mini-IPIP scales have a similar pattern of criterion-related validity as the 10-item scales. Likewise, the total amount of variance explained by the Big Five for each of the criterion variables was quite similar for the IPIP-FFM scales and the Mini-IPIP scales, as indicated by the multiple Rs.

General Discussion

Sometimes researchers need to use short assessments. This simple fact may raise concerns from reviewers of manuscripts and grant proposals and will aggravate purists who prefer the original measure to that of a shorter form. Unfortunately, there is no way

Table 6

Criterion-Related Validity for Big-Five Measures Predicting Affect and Life Satisfaction (Study 5)

Big Five Scale at Time 1	Criteria at Time 2 ($\approx$ 6–9 Months Later)		
	Positive Affect	Negative Affect	Life Satisfaction
Extraversion			
IPIP	.31	−.39	.29
Mini-IPIP	.31	−.34	.28
Agreeableness			
IPIP	.50	−.22	.27
Mini-IPIP	.43	−.09	.21
Conscientiousness			
IPIP	.25	−.14	.21
Mini-IPIP	.16	−.16	.15
Neuroticism			
IPIP	−.27	.63	−.35
Mini-IPIP	−.29	.61	−.40
Imagination/Intellect			
IPIP	.16	.01	.06
Mini-IPIP	.06	−.03	.02
R			
IPIP	.57	.67	.47
Mini-IPIP	.56	.65	.49

Note. $N = 148$.

to escape many practical constraints such as the limited patience or attention span of research participants, the fixed time period allowed for testing, and financial limits for conducting a study. To meet the practical need for short forms of broad personality traits, we attempted to create a brief measure of the Big Five in a scientifically appropriate manner by reducing the 50-item public domain IPIP-FFM inventory (Goldberg, 1999) to 20 items. The series of studies just presented give us several reasons to have confidence that we created a practically useful brief measure of the Big Five.

First, results from all five studies indicated that the Mini-IPIP scales had respectable internal consistencies given their length and content breadth. The alphas were well above .60 across all five studies. Second, the Mini-IPIP scales tapped nearly the same Big Five facet content as the IPIP-FFM scales as demonstrated in Study 2. That is, when we correlated the Mini-IPIP and the IPIP-FFM scales with a separate IPIP measure assessing the facets of the Big Five, we obtained a very similar pattern of associations. To be sure, brief scales may not capture all facets of the Big Five with equal fidelity; however, our four-item scales did not seem remarkably deficient when compared to their parent scales. The broad conclusion that we draw from Table 2 is that researchers who opt to assess the Big Five broadly can do so with relatively few items. Indeed, the TIPI scales showed a similar pattern of content coverage as the Mini-IPIP and IPIP-FFI scales. This attests to the usefulness of the TIPI for those situations where even 20 items are too many.

Moreover, the retest correlations for the Mini-IPIP scales were quite similar to the IPIP-FFM scales across intervals of a few weeks (Study 4) and several months (Study 5). Finally, the Mini-IPIP scales showed a similar pattern of convergent and criterion-related validity as compared to the 10-item IPIP-FFM scales (Studies 2–5). It is important to note that researchers might have to tolerate some reduction in validity because smaller scales have somewhat lower internal consistency reliability. Even given this possible tradeoff, the Mini-IPIP reliability coefficients were high and the validity coefficients were very similar to the IPIP-FFM scales. These pieces of evidence suggest that researchers will not sacrifice much predictive validity when using these brief scales, at least with respect to the family of criterion variables that we assessed having to do with psychological distress and well-being. On balance, any slight reduction in criterion-validity has to be weighed against the practical benefits of reducing the total number of items in the IPIP-FFM scale from 50 to 20 items (a 60% reduction).

In light of these considerations, we believe that the Mini-IPIP has much to recommend to researchers who need a brief assessment of the Big Five. Lim and Ployhart (2006) recently demonstrated that the IPIP-FFM was a very suitable replacement for the widely used NEO-FFI (Costa & McCrae, 1992) and concluded that "researchers could readily benefit from having this free alternative to the existing proprietary instruments when conducting personality-related research" (p. 50). By extension, especially in light of our study findings, it appears that the Mini-IPIP also may prove to be a very useful, efficient, and ultimately economical instrument given that it is in public domain.

In addition to creating a practically useful tool, we gained several insights into the construction of short forms that are likely to have broad applicability to researchers who opt to shorten existing measures. Foremost, we concur with Smith et al. (2000) in that researchers should begin the process of constructing short forms based on an empirically refined "parent" instrument (see also Marsh et al., 2005). The 50-item IPIP-FFM inventory is psychometrically strong in terms of the internal consistencies of its scales and its factor structure (at least from the perspective of exploratory factor analysis). These properties greatly facilitated the creation of the short form. Conversely, if the parent instrument has severe flaws, then it will be nearly impossible to create a reasonable short form from it. In addition, the content validity of the parent form places the upper limit on the content validity of the shorter form.

Second, short-form development is greatly enhanced when researchers can proceed in a theoretically informed manner. For example, by drawing on the perspective that the five factors of personality are in fact orthogonal (see Saucier, 2002), we were able to develop an additional rationale for selecting items for our short form. That is, we selected items for scales that were relatively "pure" indicators of one and only one of the Big Five. Without this basic conceptual guidance, we would have had more difficulty in the process of developing the Mini-IPIP. Moreover, a clear idea of the content span of the Big Five factors aided our selection of items, which was indicated by the distribution of the IPIP-FFM items across the facets of each Big Five factor. We have thus provided specific examples of what may be an obvious underlying principle: Developing an appropriate short form requires theoretical and conceptual understanding of the constructs under investigation.

Lastly, replication is a crucial aspect of good short form development. Indeed, Smith et al. (2000) emphasized the need for independent replication in the development of short forms, and we closely followed their guideline. We started the process with a large dataset to ensure that the factor loading matrix was reasonably stable. We then used a second independent dataset to adjust the item selection and replicated our results using three other datasets. This process increased our confidence in the Mini-IPIP, because we found that the adequacy of the short form was a repeatable phenomenon across samples and contexts. We encourage other short-form developers to rely on this method of replication when constructing their inventories. Likewise, we suggest that users of short forms demand this kind of evidence when selecting these sorts of measures for their studies.

Despite our enthusiasm for the Mini-IPIP, a few limitations and caveats should be noted. The most important issue concerns the samples that were used to develop this measure. All of our analyses were based on college-student samples, and future research should extend these results to other populations. We have no reason to suspect that the results will not generalize to other groups, but this is an empirical matter to be tested. Second, rather than examine the subset of Mini-IPIP items as we did, future research should administer the smaller scales by themselves to additional samples to examine the properties of the Mini-IPIP. We had argued that transient errors could be reduced by administering shorter forms, and thus the very brief inventory may reduce respondent fatigue and content redundancy, thereby increasing the reliability and validity of the measure compared with the results reported here. This potential benefit might be countered by the fact that participants may need more items to get in a suitable mindset

for responding to items, and in this sense fewer items may lead to greater measurement error.

Finally, there is content overlap between some of the items in the Mini-IPIP scales. This reduces the scales' construct breadth in the service of increased internal consistency. This is perhaps an inevitable consequence of our attempt to balance several potentially competing considerations: short scale length, high internal consistency reliabilities, and empirical distinctiveness between scales. As noted by Saucier and Goldberg (2002), "scale construction can serve any of many possible masters, but these masters can lead us in divergent directions" (p. 53). Although we believe that the Mini-IPIP has achieved a workable balance among these concerns, we do acknowledge that the construct breadth of the Mini-IPIP scales is a potential limitation of the measure.

All told, our studies indicate that the Mini-IPIP is a useful tool for researchers needing a very short assessment of the Big Five. Our bottom line is that the 20-item Mini-IPIP is nearly as good as the longer 50-item IPIP-FFM parent instrument in terms of both reliability and validity. This is, of course, our judgment based on the empirical findings presented here, and we stress that test users must consider the Mini-IPIP within their own context of research or application, and they should consider the many costs and benefits of using short forms when deciding whether to use the Mini-IPIP or its parent inventory. When making this decision, we believe it is useful to adopt a practical perspective, namely that of research participants who are often faced with the task of actually completing a long series of questionnaires. Our experience is that many study participants hold the view that the shorter the testing time, the better the study. We learned that it is possible to make very effective measures of broad constructs with relatively few items. As such, we suspect that many instruments might be longer than necessary and therefore could be successfully shortened by taking an approach similar to ours. In closing, we note that shortening the length of instruments may lead to subtle improvements in the experience and motivation of those participating in psychological research, one outcome that could yield big dividends for psychological science.

References

Arbuckle, J. L. (2003). *AMOS 5.0 update to the AMOS user's guide.* Chicago: Small Waters.

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.

Ball, S. A. (2005). Personality traits, problems, and disorders: Clinical applications to substance use disorders. *Journal of Research in Personality, 39,* 84–102.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187–215.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63,* 452–459.

Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment, 83,* 29–45.

Carver, C., S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67,* 319–333.

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.

Cole, J. C., Rabin, A. S., Smith, T. L., & Kaufman, A. S. (2004). Development and validation of a Rasch-derived CES-D short form. *Psychological Assessment, 16,* 360–372.

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

Costa, P. T., Jr., & Widiger, T. A. (2002). *Personality disorders and the Five-Factor Model of personality* (2nd ed.). Washington, DC: American Psychological Assocation.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffen, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49,* 71–75.

Diener, E., Smith, H., & Fujita, F. (1995). The personality structure of affect. *Journal of Personality and Social Psychology, 69,* 130–141.

Durrett, C., & Trull, T. J. (2005). An evaluation of evaluative personality terms: A comparison of the Big Seven and Five-Factor model in predicting psychopathology. *Psychological Assessment, 17,* 359–368.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286–299.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception.* San Diego: Academic Press.

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52,* 197–221.

Goldberg, L. R. (1993). The structure of personality traits: Vertical and horizontal aspects. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.). *Studying lives through time: Personality and development* (pp. 169–188). Washington, DC: American Psychological Association.

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. J. Deary, F. De Fruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37,* 504–528.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103,* 265–275.

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

Johnson, J. A. (2000). *Developing a short form of the IPIP-NEO: A report to HGW Consulting.* Unpublished manuscript. Department of Psychology, University of Pennsylvania, DuBois, PA.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Kline, R. B. (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Krueger, R. F., Caspi, A., & Moffitt, T. E. (2000). Epidemiological personology: The unifying role of personality in population-based research on problem behaviors. *Journal of Personality, 68,* 967–998.

Lim, B. C., & 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.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting

indicators for multivariate measurement and modeling latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4,* 192–211.

Lucas, R. E., Diener, E., Grob, A. Suh, E. M., & Shao, L. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology, 79,* 452–468.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100,* 107–120.

Marsh, H. W., Ellis, L. A., Parada, R. H., Richards, G., & Heubeck, B. G. (2005). A short version of the Self Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment, 17,* 81–102.

McCrae, R. R., Zonderman, A. B., Costa, P. T. Bond, M. H., & Paunonen, S. V. (1996). Evaluating the replicability of factors in the revised NEO personality inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70,* 552–566.

Miller, J. D., & Lynam, D. (2001). Structural models of personality and their relation to antisocial behavior: A meta-analytic review. *Criminology, 39,* 765–798.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89,* 187–207.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction.* New York: Harcourt College.

Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality, 35,* 463–482.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63,* 506–516.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the Big Five. *Journal of Research in Personality, 36,* 1–31.

Saucier, G., & Goldberg, L. R. (2002). Assessing the Big Five: Applications of 10 psychometric criteria to the development of marker scales. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 29–58). Seattle. WA: Hogrefe and Huber Publishers.

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods, 8,* 206–224.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89,* 623–642.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12,* 102–111.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology, 55,* 167–194.

Trull, T. J., & Sher, K. J. (1994). Relationship between the five-factor model of personality and Axis I disorders in a non-clincal sample. *Journal of Abnormal Psychology, 103,* 350–360.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76,* 820–838.

Widiger, T. A. (2005). Five factor model of personality disorder: Integrating science and practice. *Journal of Research in Personality, 39,* 67–83.

Wiggins, J. S. (2003). *Paradigms of personality assessment.* New York: Guilford Press.

Appendix

20-Item Mini-IPIP

Item	Factor	Text	Original Item Number
1	E	Am the life of the party.	1
2	A	Sympathize with others' feelings	17
3	C	Get chores done right away.	23
4	N	Have frequent mood swings.	39
5	I	Have a vivid imagination.	15
6	E	Don't talk a lot. (R)	6
7	A	Am not interested in other people's problems. (R)	22
8	C	Often forget to put things back in their proper place. (R)	28
9	N	Am relaxed most of the time. (R)	9
10	I	Am not interested in abstract ideas. (R)	20
11	E	Talk to a lot of different people at parties.	31
12	A	Feel others' emotions.	42
13	C	Like order.	33
14	N	Get upset easily.	29
15	I	Have difficulty understanding abstract ideas. (R)	10
16	E	Keep in the background. (R)	16
17	A	Am not really interested in others. (R)	32
18	C	Make a mess of things. (R)	18
19	N	Seldom feel blue. (R)	19
20	I	Do not have a good imagination. (R)	30

Note. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Neuroticism; I = Intellect/Imagination; (R) = Reverse Scored Item. Original 50-item IPIP-FFM available at http://ipip.ori.org/newQform50b5.htm.

Mini-IPIP

How much do you agree with each statement about you as you generally are now, not as you wish to be in the future?

```
1--------------------------2--------------------------3--------------------------4--------------------------5
Strongly            Somewhat            Neither Agree         Somewhat            Strongly
Disagree            Disagree            nor Disagree          Agree               Agree
```

In general, I…

1. Am the life of the party.
2. A Sympathize with others' feelings.
3. Get chores done right away.
4. Have frequent mood swings.
5. Have a vivid imagination.
6. Don't talk a lot.
7. Am not interested in other people's problems.
8. Often forget to put things back in their proper place.
9. Am relaxed most of the time.
10. I Am not interested in abstract ideas.
11. Talk to a lot of different people at parties.
12. Feel others' emotions
13. Like order.
14. Get upset easily.
15. I Have difficulty understanding abstract ideas.
16. Keep in the background.
17. Am not really interested in others.
18. Make a mess of things.
19. Seldom feel blue.
20. Do not have a good imagination.

Scoring:

Items 6, 7, 8, 9, 10, 15, 16, 17, 18, 19, and 20 need to first be reverse scored before scoring the five domains:

1=5
2=4
3=3
4=2
5=1

Neuroticism: Mean of items 4, 9r, 14, 19r
Extraversion: Mean of items 1, 6r, 11, 16r

Openness: Mean of items 5, 10r, 15r, 20r
Agreeableness: Mean of items 2, 7r, 12, 17r
Conscientiousness: Mean of items 3, 8r, 13, 18r

SPSS Syntax

```
RECODE ipip6 ipip7 ipip8 ipip9 ipip10 ipip15 ipip16 ipip17 ipip18 ipip19 ipip20 (1=5) (2=4)
(3=3) (4=2) (5=1) INTO ipip6r ipip7r ipip8r ipip9r ipip10r ipip15r ipip16r ipip17r ipip18r
ipip19r ipip20r.
EXECUTE.

COMPUTE Neuroticism=MEAN(ipip4,ipip9r,ipip14,ipip19r).
COMPUTE Extraversion=MEAN(ipip1,ipip6r,ipip11,ipip16r).
COMPUTE Openness=MEAN(ipip5,ipip10r,ipip15r,ipip20r).
COMPUTE Agreeableness=MEAN(ipip2,ipip7r,ipip12,ipip17r).
COMPUTE Conscientiousness=MEAN(ipip3,ipip8r,ipip13,ipip18r).
EXECUTE.
```