Journal of Research in Personality **36,** 1–31 (2002)
doi:10.1006/jrpe.2001.2335, available online at http://www.idealibrary.com on IDEAL®

Orthogonal Markers for Orthogonal Factors: The Case of the Big Five

Gerard Saucier

University of Oregon

The Big Five model, originally based on studies of lexical personality descriptors, has proved useful and attracted much consensus. However, although the Big Five factors are held to be orthogonal, they are not orthogonal in previous marker sets developed for them. The author describes methods for creating relatively orthogonal marker sets for orthogonal factors and demonstrates that the Big Five can be operationalized in relatively orthogonal marker scales. The three marker sets presented are a 40-item short form of Goldberg's 100 unipolar markers (Ortho-Markers), a new set of Modular Markers built with item parcels, and a 40-item short form of these (Mini-Modular Markers [the 3M40]). The new Big Five marker sets are shown to have markedly lower interscale correlations, with no loss of validity, relative to previous marker sets with comparable numbers of items. This indicates that the nonorthogonality of previous Big Five marker scales is less a property of the Big Five model than an unintended outcome of commonly used scale construction procedures. © 2002 Elsevier Science (USA)

The past decade has seen a great burst of interest in the search for a scientifically compelling taxonomy of personality attributes. Such a taxonomy will help to organize and integrate knowledge and research findings by providing a standard scientific nomenclature to facilitate communication among investigators and to aid in the accumulation of empirical findings.

The beginning of a consensus has emerged about the general framework of such a taxonomic representation. One structural model (the Big Five) has become increasingly popular, but it is not yet clear that this is the "optimal" model. An optimal model will be replicable across methods, cross-culturally generalizable, comprehensive, and high in utility. Although there are some questions about its cross-cultural generalizability, the utility of the Big Five has been amply demonstrated, its comprehensiveness with respect to the con-

Work on this article was supported by Grant MH-49227 from the National Institute of Mental Health, U.S. Public Health Service.

Address correspondence and reprint requests to Gerard Saucier, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403. E-mail: gsaucier@oregon.uoregon.edu.

0092-6566/02 $35.00

tent of most personality inventories appears good, and its replicability across methods (at least in the American context) seems good.

Measurement in the Big Five paradigm is currently beset by one particular discrepancy between theory and practice. The Big Five were discovered—repeatedly, in fact, by a variety of investigators (for reviews, see Digman, 1990; John, 1990)—in analyses using orthogonal rotations, and variation on each one of the Big Five dimensions is commonly proposed to be independent of variation on each of the others. In other words, the Big Five are "supposed" to be orthogonal or have been orthogonal in theory. In practice, however, the scales used to measure them are not mutually orthogonal. When measures of a model do not measure what they are "supposed" to measure, when the "talk" does not match the "walk," construct validity is at issue. Correlations among dimensions, although not typically problematic, are a problem when one has a model that posits their nonexistence. In this article, I address this problematic discrepancy by reporting attempts to develop relatively orthogonal marker sets for the Big Five. If such marker sets can be developed, one might eliminate this discrepancy between the talk and the walk associated with the Big Five.

ARE THE BIG FIVE FACTORS TRULY ORTHOGONAL?

Block's (1995; see also Funder, 2001) critique of the "five-factor approach" noted that "the empirical research findings indicate that the five factors are frequently importantly correlated with each other, usually to reflect an overriding evaluative component" (p. 199), and that "the failure of orthogonality is not a one-time ephemeral finding" (p. 200) but rather a regular occurrence. To be clear, Block was not discussing factors but rather was discussing scales designed to measure the factors. More specifically, he was referring to interscale correlations reported in published sources. For example, in self-ratings, Goldberg's (1992) unipolar markers were reported to have intercorrelations as high as .37. And high interscale correlations are certainly not confined to lexical studies or adjective stimuli. The Revised NEO Personality Inventory (NEO-PI-R) (Costa & McCrae, 1992) has reported domain–scale intercorrelations of magnitude as great as .53 in self-ratings. Another questionnaire, the Big Five Inventory (BFI) (Benet-Martínez & John, 1998), has some reported interscale correlations—in both English and Spanish versions—that exceed .30 in self-ratings. The average reported scale intercorrelations for each of these three instruments is about .20, but these averages are misleading in one respect: Intercorrelations of specific pairs of factors are often quite high, and one such high intercorrelation is enough to call into question the assumption of "five orthogonal factors."

Block (1995) noted that the claim of "nearly orthogonal" factors was best realized when data are "restricted to self-descriptions or the descriptions of

liked others are analyzed" (p. 199) but that "when data based on ratings of a subject pool including disliked individuals are factored, the five-factor markers issue a factor structure that is impressively nonorthogonal" (p. 200). Block was referring to a large number of scale intercorrelations exceeding .30, and as high as .58, reported for Goldberg's 100 Markers (Goldberg, 1992, Table 5) in such a heterogeneous set of data. Reports using other Big Five measures have focused on data including self-reports and characterizations of liked peers and intimates. However, samples in personality research should reflect the reality that some persons are judged to have undesirable attributes. That is, the "study of personality differences cannot be limited to data from restricted, homogeneous samples of laypersons describing themselves and their friends" (Block, 1995, p. 200).

If the Big Five factors are actually orthogonal, this would be an advantageous feature of the Big Five model for two reasons. First, when one is mapping a domain of variables, as when one is mapping a physical landscape, orthogonal axes provided a superior coordinate system for locating points on the map. As a result, those psychometric theorists (e.g., Guilford, 1954) who have emphasized the use of factors to represent tests graphically have usually preferred orthogonal rotations. Second, orthogonal predictors are more efficient in multiple-regression analyses because they minimize multicollinearity and maximize discriminant validity; as Jackson (1971) noted, "If one wishes to maximize the predictability of a battery, entirely uncorrelated tests would be appropriate" (p. 246). Because the Big Five are so widely used as a taxonomic coordinate system for personality variables, and as predictors in applied research, orthogonal factors are crucial to the model. If the Big Five are *not* orthogonal, they will be less useful as a structural map and less efficient as a set of predictors.

Of course, orthogonal factor models are not always (and, one might argue, not usually) superior to oblique factor models. But orthogonal factor models have distinct virtues, and these virtues have been an important part of the Big Five concept. To be valid, any scientific model should be described and labeled as accurately as possible; an orthogonal factors model that is not what it is claimed to be can lead to inferential errors when the model is applied in real-world situations.

Recognition of nonorthogonality in marker scales has prompted some statistical remedies, such as (a) the ipsatization of data, which tends to lower the scale intercorrelations (Goldberg, 1990, 1992); (b) the use of orthogonal (e.g., varimax) factor scores (Goldberg, 1992); and (c) the generation of the desired factors via factor score coefficients or orthogonal Procrustes rotations (Costa & McCrae, 1992; McCrae et al., 1996). These remedies are arguably problematic.

Ipsatization (making the summed scores for each individual the same, e.g., by standardizing them) lowers Big Five interscale correlations because it

removes some variance attributable to a general factor, which is largely responsible for these unwanted correlations. Ipsatization also controls for acquiescence and extremeness (vs moderacy) responding.[1] But ipsatizing is a controversial remedy because equalizing subject means (to 0) and equalizing subject variances (to 1) might not be justified; some subjects may have more of the traits referenced in the items, or more extreme traits, than do other subjects, but ipsatizing effaces these differences. Ipsatizing (or otherwise controlling for acquiescence) across sets of items that do not have balanced keying with respect to content can lead to inadvertently discarding content variance (Hofstee, 1998). And in ipsatizing, researchers control for between-subjects differences in spread (variance) as well as central tendency (mean), often with no clear rationale for doing so (Hofstee, 1998).

Scoring subjects using factor scores has its own problems. When based on exploratory factor analyses, these factor scores are calculated using new weights in each new sample and thus are less uniform across samples than are scale scores. And if one imposes old (previously used) weights on a new sample, these weights are prone not to generate the same degree of orthogonality as in the sample in which the weights were derived. Thus, in one way or the other, factor score weights suffer from sample-dependence. Moreover, factor score weights and ipsatization share an important practical limitation. Most researchers and users of Big Five markers prefer a more straightforward scoring approach and continue to make use of simple (but correlated) scale scores based on raw data, oblivious to these statistical remedies. Accordingly, a close approximation to orthogonality would be a desirable and practical feature in a Big Five marker set, unless the premise of mutual orthogonality is simply abandoned.

Should the premise indeed be abandoned? Perhaps the clusters of variables the five factors represent are fundamentally oblique, so that they would be more accurately represented by a correlated factors model. For example, Digman (1997) derived recurrent higher order factors from intercorrelations among five-factor scales, in effect treating the five factors as oblique first-order factors, even though that author elsewhere regarded the Big Five as five orthogonal factors. If the Big Five are fundamentally oblique, then mutually orthogonal scales will either (a) be impossible to create, (b) have markedly

[1] In personality data, the general factor is most usually interpretable in large part as a social desirability (good vs bad) dimension. In broad heterogeneous selections of personality variables, the general factor is typically not as salient as in some other domains of variables (e.g., cognitive ability measures), so the effect of ipsatizing is less profound than in many other domains. The classic objections of Clemans (1966) to ipsatized data are considerably weakened when the average correlation of variables is not positive but rather close to zero (ten Berge, 1999). Nonetheless, a correlation matrix of ipsatized personality data still tends to have a more negative average correlation than does that derived from the original (raw) data, resulting in a greater tendency toward bipolar factors—factors with items loading on both poles.

lower validity, or (c) transform the content of one or more of the factors into something unfamiliar, not interpretable in the same way as before. If the latter (c) occurred, this would suggest that the factors are meaningful when mutually orthogonal but have been previously interpreted (perhaps misunderstood) in terms of constructs that are inherently correlated. Thus, an attempt to create mutually orthogonal marker scales can be revealing with respect to the nature of the Big Five factors.

Why might fundamentally orthogonal factors tend to be operationalized in obliquely related scales? Overly expansive conceptualizations may be at fault. Many assume that the Big Five all are equally big. But it may be that, as Peabody and Goldberg's (1989) results indicate, three of the Big Five factors (Extroversion, Agreeableness, and Conscientiousness) are broad and the other two (Emotional Stability and Intellect) are not so broad. If so, then attempts to broaden the latter two factors would do so largely by taking up content that could be subsumed under the first three factors; the redundancy would create higher correlations between the latter factors and the first three factors. For example, in extant scales, Agreeableness (II+) and Emotional Stability (IV+) tend to be positively correlated. As representations of a circumplex defined by Factors II and IV (Hofstee, De Raad, & Goldberg, 1992; Saucier, 1992) indicate, there are many variables in the II+/IV+ and II−/IV− interstitial areas, that is, many variables for "stable agreeableness" and "disagreeable neuroticism" and fewer for stable disagreeableness and agreeable neuroticism. One way to make the Neuroticism (IV−) factor larger would be to include and emphasize hostility/irritability-related content in this factor, although this content is also highly related to (low) Agreeableness. Such a step would in effect push the centroid of the Emotional Stability (IV) scale in the direction of Agreeableness (II) and increase the correlation between these two marker scales.

Or perhaps the nonorthogonality of extant Big Five scales is an expectable outcome given the most commonly used scale construction procedures. One typical approach is to construct a marker scale for a factor by simply adopting those items in the item pool that demonstrate the highest loadings on a given factor. This method is valuable because it provides a set of items that best define a factor while also tending to be mutually correlated. Thus, it tends to maximize reliability; indeed, another typical approach is to select that set of items that best maximizes coefficient alpha (internal consistency) reliability. These two related approaches have several important limitations. Choosing only highest loading items, or items that together maximize alpha, does not guarantee interscale orthogonality (if that is important), nor does it maximize breadth. The breadth of content reference possessed by the scale importantly affects its validity with respect to a *broad* range of criteria (although not necessarily to some narrow range of criteria). Scales that singularly maximize reliability may fall prey to elements of the "attenuation para-

dox'' (Loevinger, 1954): a decrease in validity as one increases reliability by restricting a scale to a narrowed range of content. To prevent this restriction of validity, an ideal marker scale would have its correlation with the criterion factor maximized. But such validity maximization may dictate inclusion of content not found among the very highest loading items. The goal of ensuring orthogonality also may dictate inclusion of more moderately loading items. And making the scale more equidiscriminating (Nunnally & Bernstein, 1994)—able to measure well at all levels of the underlying dimension—is also likely to mean going beyond the highest loading items. In general, item selection strategies that dictate inclusion of only the highest loading items, or generally focus on maximizing reliability, may generate less orthogonal marker scales than might strategies emphasizing other criteria.

In summary, it has not been clear whether the Big Five can be operationalized in mutually orthogonal marker scales. And if mutually orthogonal scales can indeed be created, we know little about whether such scales would need to be interpreted and labeled differently from previous Big Five scales, in other words, whether they would be representing the same Big Five.

Attempts to generate mutually orthogonal scales could be quite informative with respects to these important issues. I detail attempts based on (a) the item pool provided by a widely used Big Five marker set and (b) a larger, more representative item pool, with aggregation of items into parcels and some attention to equidiscriminating criteria.

AN ORTHOGONAL SUBSET OF GOLDBERG'S ADJECTIVE MARKERS

Goldberg (1992) presented a set of 100 adjectives that can be scored (20 items per factor) to yield the Big Five; indeed, Block (1995) referenced this marker set extensively in his critique of the Big Five with regard to orthogonality. This is a widely used marker set that can be found in many researchers' archival data; a subset of them scored to yield orthogonal scales would certainly be useful.

From a pool of 566 reasonably common personality-descriptive adjectives, Goldberg (1992) selected 100 terms by selecting those having high loadings on the given factor and low loadings on other factors and then employing criteria of (a) internal consistency and (b) replicability in factor loading patterns across three samples of subjects. To make the factors relatively equal in size, 20 terms were selected for each factor.[2] To reduce the effect of individual differences in response scale use, 10 items were selected for the positive and 10 for the negative pole of each factor, with the exception of Factor IV (Emotional Stability), where a dearth of suitable positive items led to a

[2] If one's goal was an even more complete equality of the scale variances, one might have had to select slightly different numbers of items for each of the five factors.

mix of 6 positive and 14 negative items. Goldberg showed that each of the five 20-item subsets of these 100 Markers, when considered as separate scales, yielded highly reliable scores; the mean alpha coefficients ranged from .85 to .93 depending on the data set, with all of the individual coefficients above .80 in each data set.

I scrutinized these 100 Marker items in the same self- and peer-rating data sets used in Goldberg's (1992) Study 4. Sample A included ratings by 316 university students of peers of the same sex that they knew well and liked. Sample B included 205 ratings of same-sex peers, with approximately one-third each being liked, disliked, and neutrally regarded targets. Sample C included self-ratings by 320 university students, and Sample D included self-ratings by a separate set of 187 university students. Details concerning these samples are found in the upper portion of Table 1.

The items that contribute most to the positive scale intercorrelations can be identified most readily by examining the secondary loadings of the items. These problem items were removed until 8 items remained per scale (while maintaining balanced keying as much as feasible). This 40-item subset is labeled the Ortho-40 and is presented in Table 2. Coefficient alpha reliability averages about .10 lower than for the 100-marker scales. But interscale correlations are dramatically lower than for the full set of 100 Markers, on average about .15 lower per pair of scales.[3] The highest interscale correlations are in the Pooled Peer sample, where the general evaluation factor has a powerful effect on these coefficients; in this extreme case, whereas one correlation in the 100 Markers reached .58 (Factors II and IV), the highest correlation in the Ortho-40 was .40 (Factors II and V). Overall, the Ortho-40 sacrifices some reliability to gain greater mutual orthogonality.

[3] Means of the absolute values of interscale correlations were .05, .08, .19, and .08 in the Self, Liked Peer, Pooled Peer, and Community samples, respectively. One might assume this mean of absolute values to be a superior index of mutual orthogonality, but this index is sensitive to sample size, especially for the most mutually orthogonal variables. For a set of correlations whose population (true) values approach .00, as sample size decreases (especially below $N = 100$), the absolute values of sample (obtained) correlations become more extreme (highly positive or highly negative), so that their absolute values are inflated by increased contributions of chance variation. Thus, the means of the absolute values of these correlations will reflect this inflation. By contrast, if the true correlations in the population are instead about .20, the absolute values of obtained correlations show much less sensitivity to sample size because they will always be predominantly of the same sign (positive) rather than of mixed signs as in the first case; chance variation will as often as not make obtained correlations lower (in absolute value) than the true correlations. The mean of obtained (not absolute) values of interscale correlations, which are reported in the table, are comparatively invariant with changes in sample size; specifically, they do not become inflated at smaller sample sizes. And they directly indicate the degree to which a set of scales all scored in the desirable (or scored in the undesirable) direction do or do not yield a positive manifold reflecting the overriding evaluative component to which Block (1995) referred, long known to be an important element in personality ratings.

TABLE 1
Characteristics of Samples

Label	Stimulus set (number of adjectives)	*N*	Percentage women	Characteristics of raters	Targets of description	Year of data collection
A	587	316	62	University undergraduates	Liked peer of same sex	1978
B	566	205	41	University undergraduate, graduate psychology, and law students	Same-sex peer[a]: 1/3 liked, 1/3 disliked, 1/3 neutral	1976
C	587	320	63	University undergraduates	Self	1978
D	1,710	187	63	University undergraduates	Self	1972
E	791[b]	700	57[c]	Community residents (born 1914–1974)[e]	Self	1993–1998
		592	58			
		394	57[d]			
F	600	201	72	Community college students and university undergraduates	Liked peer of same sex	1995

Note. Further information on Samples A through D can be found in Goldberg (1982).

[a] Individual participants were randomly assigned to one of these three kinds of targets.

[b] Sample E was administered 360 trait-descriptive adjectives in 1993, 525 person-descriptive adjectives in 1995, and 216 in 1998, for a total of 791 distinct terms, with administration of other inventories (NEO-PI-R, NEO-FFI, BFI, act frequency ratings) at other points in the 1993–1998 time span.

[c] Subsample of 592 that could be scored for all Modular Marker scales.

[d] Subsample of 394 participants who completed all measures used in comparative validity analyses.

[e] Mean age of sample in 1993 was approximately 51 years (SD = 12).

TABLE 2
Orthogonal Subset of 100 Unipolar Markets (Ortho-40): Reliability and Interscale Correlations

	Derivation samples			Cross-validation sample
	Self	Liked Peer	Pooled Peer	Community Sample
Coefficient alpha				
I	.84	.86	.86	.81
II	.73	.79	.92	.71
III	.86	.87	.89	.85
IV	.70	.62	.70	.72
V	.71	.72	.83	.74
Interscale correlations				
I–II	.03	.02	−.10	.03
I–III	−.07	−.12	−.08	.04
I–IV	.00	.02	−.17	.05
I–V	.03	.06	−.13	.10
II–III	.11	.15	.19	.12
II–IV	−.05	.03	.30	.19
II–V	.04	.21	.40	.06
III–IV	−.03	.09	.14	.13
III–V	.08	.06	.28	.06
IV–V	−.06	−.04	.10	−.03
Ortho-40 mean	.01	.05	.09	.07
100 Markers mean	.13	.24	.27	.25
Mini-Markers mean	.11	.18	.26	.22

Note. Self sample N = 320, Liked Peer sample N = 316, Pooled Peer sample N = 205, Community Sample, N = 1,125. All analyses were conducted with raw (non-ipsatized) data. Scale items are as follows: I—Bold, Extraverted, Talkative, Unrestrained versus Introverted, Quiet, Reserved, Shy; II—Kind, Sympathetic, Undemanding, Warm versus Cold, Demanding, Harsh, Unsympathetic; III—Efficient, Neat, Organized, Systematic versus Careless, Disorganized, Sloppy, Unsystematic; IV—Unenvious, Unexcitable versus Anxious, Emotional, Fearful, Fretful, Nervous, Touchy; V—Artistic, Complex, Creative, Deep, Introspective, Philosophical versus Simple, Unreflective.

The right-most column in Table 2 shows reliability and interscale correlations in a new "cross-validation" sample, Sample E. This is a community sample whose members were recruited by and who participate by mail. Cross-validation is important because one cannot expect scales to have uniform levels of intercorrelation, or an equal degree of mutual orthogonality, in all samples; indeed, such orthogonality in the derivation sample(s) could be due to idiosyncrasies of the sample(s). In this additional sample, interscale correlations were also low (averaging .07), indicating that the near orthogonality of the Ortho-40 scales is not a strongly sample-dependent phenomenon.

Does making the Big Five more mutually orthogonal notably affect their content? Such an effect, if any, appears to be small. Ortho-40 items for Extroversion, Agreeableness, and Conscientiousness differ little from those in previous marker sets. Emotional Stability appears to emphasize Anxiety and deemphasize Irritability more than do previous marker sets. The Intellect factor appears to emphasize perceptions of intellectual *ability* less than do some previous marker sets; 100 Marker terms such as Bright and Unintelligent are omitted. Instead, there is more emphasis on an intellectualist *style* (e.g., Philosophical, Introspective, Deep). But these shifts are subtle and do not seem to fundamentally alter the character of the factors.

The Ortho-40 subset provides one demonstration that the Big Five are not oblique by necessity. If one's item pool permits, it should be possible to develop a set of marker scales that are virtually uncorrelated.

However, the Ortho-40 is not an optimal Big Five marker set in several ways. As noted, the reliability of the scales, although falling in an acceptable range, are lower than some might like. One of the interscale correlations reached as high as .40. Moreover, this marker set can be scored for five factors but not for more specific subcomponents of these factors; that is, it cannot be systematically decomposed into parcels. And there is a question as to whether the item pool of the 100 Markers is an ideal starting point; perhaps starting from a more representative item pool would allow superior content representation. In search of a less imperfect solution to the orthogonal factor markers problems, I used a different strategy.

A MODULAR MARKER SET

My second attempt at creating mutually orthogonal scales was based on a more complex psychometric strategy than was the first. I begin by detailing the bases of this more complex approach, which includes three emphases absent from the Ortho-40.

Representative Sampling of Variables

Measures of the Big Five are sometimes misconstrued as a representation of all important personality variables. This misconstrual is based on a failure to differentiate between "representative sets of variables" and "marker sets." Goldberg (1992) distinguished between "representative sampling of the total domain" and "cluster sampling, which aims for factor-univocal variables by the systematic omission of those located in interstitial regions between the clusters" (p. 28).[4] Marker sets for the Big Five (Goldberg, 1992; Benet-Martínez & John, 1998; Saucier, 1994a) are an application of cluster

[4] A third sampling strategy described by Goldberg is "uniform sampling"; this approach is used in developing markers for a circumplex (e.g., Saucier, Ostendorf, & Peabody, 2001; Wiggins, 1979) and is outside the scope of this article.

sampling, which advantageously maximizes the reliability and brevity of a measure but does so by systematically omitting variables that either (a) load substantially on two or more factors or (b) do not load very highly on any factors. Representative sampling, by contrast, attempts to include representatives of as wide a range of variables as possible. Representative sampling does not maximize the reliability or brevity of measures but does the best job of encapsulating an entire domain of variables and often leads to a measure that is associated with a wider range of criterion variables (Goldberg, 1972; Saucier & Goldberg, in press a). The Ortho-40 is based on Goldberg's 100 Markers, and therefore, because these 100 items were selected based on cluster sampling rather than representatively sampling, the Ortho-40 may be based on an overly restricted item pool. Orthogonal factor markers based on a representative set might look different.

Two representative sets of personality variables are prominent in the research literature. Peabody (1987) reduced a set of 571 personality adjectives, identified based on previous work (e.g., Goldberg, 1982; Norman, 1967), to a set of 53 bipolar pairs. Goldberg's (1990) set of 100 clusters is another attempt to representatively sample the domain, although the clusters are overtly grouped into Big Five supercategories. I chose not to employ these representative sets here because (a) both classifications omitted terms that did not clearly appear to be personality dispositions, with some potential for arbitrary omission; (b) the exclusion criteria used as a rationale for omissions were not clearly spelled out; (c) both sets were based on a larger pool of terms that are not necessarily those of most salience (i.e., high frequency of use) in the language; and (d) the procedures used to form pairs and clusters were not as precisely specified and repeatable as one might wish.

My overall strategy involved creating mutually orthogonal Big Five marker scales beginning with a representative sampling of variables. This representative sampling involved reducing to a smaller set of variables that set of 500 person-descriptive English adjectives that were determined, in a previous study (Saucier, 1997), to be of highest frequency of use, based on highly reliable, aggregated ratings of groups of judges. All 500 of these terms were used as stimuli for self-ratings from the Eugene–Springfield community sample (Sample E, $N = 700$), including adults ages about 20 to 90 years, recruited by mail. The same 500 terms were used as stimuli for ratings of liked peers by college and community college students (Sample F, $N = 201$). Data from both Samples E and F were analyzed with respect to replicability of factor structures in an earlier study (Saucier, 1997), and I combined them for these analyses, yielding Sample EF.

Parcels

Use of parcels enables one to decompose scores from a broad factor into reliable subcomponents. Moreover, it helps to ensure the meaningfulness of

broad factors because item-level factor analysis "often results in extremely specific, sometimes even trivial, exemplars of a trait" (Jackson, 1971, p. 244). And Cattell (1957) suggested that "a single item response is too chancy an event to build a complex 'pattern calculation' upon" (p. 170); single items are too prone to be interpreted differently by different respondents or by the same respondent on different occasions, and they are too prone to be affected by momentary error and by the idiosyncratic context provided by nearby items.

The 500 terms were clustered by an explicit and repeatable algorithm that tended to maximize the number of mutually discriminated, high-reliability aggregates of variables that can be extracted from an item pool. All of the coefficients in the 500 × 500 matrix of correlations in the item pool (based on original, not ipsatized, data from Sample EF) were rank ordered from highest to lowest. If each of a pair of terms had its highest positive or highest negative correlation with the other term in the pair—if the terms formed a "mutually highest correlating pair"—and that correlation exceeded a minimum magnitude (+.40 for positive correlations, −.30 for negative correlations), it was retained as the nucleus for a representative parcel. A lower criterion value was used for negative correlations because (a) bipolar pairs are more desirable from the standpoint of balanced item keying and (b) in raw data, individual differences in acquiescence tend to make the average correlation slightly positive and tend to make high positive correlations somewhat more probable than high negative ones.

After the nuclear pair for a representative parcel was identified, more terms were added, to the extent that the item pool contained sufficient items for doing so, if needed (a) to raise the coefficient alpha for the parcel to approximately .70 or higher or (b) to create more balanced keying. Because Sample E had data available using an additional 291 adjectives administered at other times, this procedure was carried out strictly in that sample. As an example of the procedure, one representative parcel was based on the mutually highest correlating pair Jealous–Possessive. To increase reliability and introduce reverse-keyed items, the item pool was searched and the additional terms Envious and Unenvious were added to yield a 4-item parcel with a reliability of .76. The 100 representative parcels formed by this procedure are detailed in Table 3. Note that the last 30 or so parcels include content not conventionally thought of as relating to personality (e.g., Wealth, Popularity, Attractiveness, Tallness) but nonetheless found among the 500 highest frequency person-descriptive adjectives; as will be seen, none of these "non-personality" parcels made its way into Big Five marker sets.

These parcels are an example of representative sampling of a domain of variables. But a set of factor markers should represent factors, not a whole domain. Variables that have a weak or overly complex relation to the factors, then, must be omitted to create a marker set. The range of content in the

TABLE 3
100 Representative Parcels and 17 Supplementary Parcels

Label, constituent terms, and coefficient alpha reliability	Scoring
Spontaneity: Impulsive, Spontaneous, *Playful* (.65, .59)	I
Loudness: Loud, Noisy (.79, —)	
Talkativeness: Talkative, (−) Quiet (.61, .77)	I
Humor: Humorous, Witty, (−) *Humorless* (.73[a], .72)	
Shyness: Shy, Bashful (.83, .81)	I−
Sociability: Sociable, (−) *Unsociable,* (−) *Withdrawn* (.73[b], .78)	I
Enthusiasm: Eager, Enthusiastic (.68, .69[c])	
Energy: Active, Energetic, (−) *Unenergetic* (.78[b], .80)	
Tiredness: Tired, Exhausted (.77, —)	
Directness: Direct, Straightforward (.73, .74)	
Assertiveness: Dominant, *Assertive, Forceful,* (−) *Timid* [Aggressive] (.68[a], .75)	I
Strength: Strong, (−) Weak, (−) *Helpless* (.63, .75[d])	
Courage: Brave, Courageous (.78, .76)	
Adventurousness: Daring, Adventurous, (−) *Unadventurous* (.77[a], .78)	I
Confidence: Confident, (−) Insecure, *Self-Assured* (.77, .72)	
Optimism: Optimistic, (−) Bitter, (−) Negativistic (.65[b] .74)	
Easygoing: Easygoing, Carefree, *Happy-go-lucky* (.66[b] .75)	
Calmness: Relaxed, *Calm,* (−) *Tense* [(−) Nervous] (.69, —)	
Fear: Afraid, Scared (.77, —)	
Self-Abasement: Ashamed, Humiliated, *Embarrassed, Guilty* (.74, —)	
Jealousy/Envy: Jealous, Possessive, *Envious,* (−) *Unenvious* (.67, .76)	IV−
Moodiness: Moody, Temperamental, Irritable (.77, .74)	
Stability: Stable, (−) Unstable, Well-adjusted, (−) *Confused* (.80, —)	
Irritation: Annoyed, Irritated (.74, —)	
Hot temper: Hot-tempered, Short-tempered (.80, —)	
Patience: Patient, (−) Impatient, (.63, .70)	
Gratitude: Grateful, Thankful (.84, —)	
Happiness: Joyful, Happy, (−) Unhappy [Cheerful] (.76, —)	
Sadness: Sad, *Depressed* [(−) Happy] (.74, —)	
Friendliness: Friendly, (−) Unfriendly (.66, .73)	
Warmth: Warm, (−) Cold (.64, .73)	II
Sympathy: Sympathetic, *Compassionate* [Understanding] (.75, .78)	II
Sensitivity: Sensitive, Sentimental [(−) Insensitive] (.48, .66)	II
Affection/Passion: Passionate, Romantic, (−) *Unaffectionate* (.70[b], .79[d])	
Toughness: Rough, Tough, *Stern, Harsh* (.58[b], .73)	II−
Kindness: (−) Cruel, *Kind* [Abusive] (.56, .66)	II
Generosity: Generous, (−) Stingy, (−) *Greedy,* (−) *Selfish* (.73, .79)	
Snobbishness: Snobbish, Stuck-up (.81, —)	
Modesty: (−) Conceited, (−) *Boastful, Modest, Humble* [(−) Egotistical] (.58[b], .82)	
Consideration: Considerate, (−) Inconsiderate (.68, .77)	
Respectfulness: Respectful, (−) Disrespectful (.65, .74)	
Politeness: Polite, (−) Impolite, Courteous (.71, .85)	
Offensiveness: Insulting, *Offensive* [Rude] (.72, —)	
Bothersomeness: Annoying, Aggravating (.78, —)	
Fairness: Reasonable, (−) Unreasonable, *Fair,* (−) *Unfair* (.74, —)	
Phoniness: Phony, Artificial (.72, —)	

Errata: Two parcels inadvertently omitted from Table 3 are–
Independence: Independent, Self-sufficient (.70, –)
Grumpiness: Crabby, Grumpy (.84, .77)

TABLE 3—*Continued*

Label, constituent terms, and coefficient alpha reliability	Scoring
Honesty: Honest, (−) Dishonest, Truthful (.69, .83)	
Faithfulness: Faithful, (−) Unfaithful (.73, —)	
Responsibility: Responsible, (−) Irresponsible (.73, —)	
Dependability: Dependable, Reliable (.81, .83)	
Promptness: Prompt, Punctual (.87, .87)	
Organization: Organized, (−) Disorganized (.80, .82)	III
Neatness: Neat, (−) Sloppy (.78, .81[d])	
Strictness: Strict, Firm (.60, .47)	
Dedication: Dedicated, Hard-working, *Determined* (.74, —)	
Caution: Careful, Cautious (.70, .77)	III
Predictability: Predictable, (−) Unpredictable (.61, .74)	
Practicality: Practical, Sensible (.76, —)	
Alertness: Awake, Alert (.67, —)	
Perceptiveness: Perceptive, (−) Shallow, *Wise*, (−) *Shortsighted* (.58[a], .74)	
Competence: Competent, (−) Incompetent, Capable (.72, —)	
Talent: Gifted, Talented (.77, —)	
Intelligence: Intelligent, Smart (.81, .86)	
Imagination: Creative, Imaginative, (−) *Unimaginative* (.78[a], .83)	
Open-mindedness: Open-minded, (−) Narrow-minded [(−) Closed-minded] (.65, 58[e])	
Liberalism: Liberal, (−) Conservative (.70, .58)	
Unconventionality: (−) Traditional, (−) *Conventional, Unconventional, Nonconforming, Rebellious* [(−) Old-fashioned] (.76, .74)	
Femininity (vs Masculinity): Feminine, (−) Masculine (.84, .89)	
Youthfulness: Youthful, Young, (−) Elderly, (−) Old (.81, —)	
Maturity: Mature, (−) Immature (.64, .75)	
Employment: Employed, (−) Unemployed (.79, —)	
Accomplishment: Accomplished, Successful, *Skilled* (.76, —)	
Wealth: Rich, Wealthy (.82, —)	
Education: Educated, (−) Uneducated (.72, —)	
Luck: Lucky, (−) Unlucky (.71, —)	
Influence: Powerful, Influential (.71, —)	
Fame: Famous, Well-known (.67, —)	
Neglect: Negleted, (−) Appreciated, *Lonesome* (.65, —)	
Likeability: Likeable, *Lovable, Pleasing* [Well-liked] (.75, —)	
Popularity: Popular, (−) Unpopular (.68, —)	
Worth: Worthy, Valuable, *Respected* (.70, —)	
Superlativeness: Great, Terrific, *Wonderful* (.86, —)	
Exceptionality: Extraordinary, (−) Ordinary, Exceptional, (−) Average (.77, —)	
Strangeness: Weird, Strange (.84, —)	
Privation: Homeless, Senile, *Blind* (.66, —)	
Stupidity: Dumb, Stupid (.74, —)	
Stimulus Value: Interesting, Fascinating, (−) Boring (.70, —)	
Appeal: Appealing, Desirable (.78, —)	
Attractiveness: Attractive, (−) Unattractive (.72, .83[d])	
Beauty: Pretty, Beautiful (.83, —)	
Sexiness: Sexy, *Sensual, Flirtatious* [Seductive] (.62[c], .70)	

TABLE 3—*Continued*

Label, constituent terms, and coefficient alpha reliability	Scoring
Awkwardness: Ungraceful, (−) Graceful, Awkward, Clumsy (.72, —)	
Athleticism: Athletic, Muscular (.71, —)	
Slenderness: Slender, (−) Chubby (.81, —)	
Tallness: Tall, (−) Short (.80, —)	
Tininess: Tiny, Little (.74, —)	
Healthiness: Healthy, Well (.79, —)	
Disability: Disabled, Handicapped (.83, —)	
Supplementary parcels employed in Big Five marker scales:	
Restraint: Inhibited, Reserved, Restrained (.52[c], .62)	I−
Agreeableness: Agreeable, Tolerant, Lenient (.52[d], .65)	II
Criticalness: Critical, (−) Uncritical (.44[a], .63)	II−
Demandingness: Demanding, (−) Undemanding (.53[a], .70)	II−
Slyness: Sly, Cunning, Shrewd (.59[c], .69)	II−
Ambition: Ambitious, (−) Unambitious (.78, .71)	III
Decisiveness: Decisive, (−) Indecisive (.58[b], .66)	III
Efficiency: Efficient, (−) Inefficient, (−) Careless (.70, .69)	III
Perfectionism: Perfectionistic, Exacting, Meticulous, Precise (.74[c], .75)	III
Forgetfulness: Forgetful, Absent-minded, Scatterbrained (.73[b], .76)	III−
Emotional Excitability: Excitable, Emotional, (−) Unexcitable (.63[a], .72)	IV−
Anxiety: Anxious, Nervous, High-strung (.73[a], .63)	IV−
Fretfulness: Fretful, Fearful (.43[a], .54)	IV−
Hyperdevotedness: Overloyal, Overprotective, Overconscientious, Oversentimental (.70, .61[d])	IV−
Intellectuality: Intellectual, (−) Unintellectual (.70, .71)	V
Analytical Inquiry: Philosophical, Deep, Complex, Analytical (.67[b], .70)	V
Reflectiveness: Introspective, Contemplative, (−) Unreflective (.57[a], .73)	V

Note. ESPS, Eugene–Springfield Community Sample combined with college peer-rating sample, $N = 901$; ABCD, combined college student Samples A, B, C, and D, $N = 1028$. Coefficients in parentheses are, respectively, coefficient alphas in ESPS and ABCD. Subscript letters indicate sample size for all items in parcel: [a]$N = 694$, [b]$N = 596$, [c]$N = 592$, [d]$N = 841$, [e]$N = 823$; — (—) indicates all of the items in the parcel not available in data from Sample A, B, C, or D. Terms in brackets were original core pair constituents that were later removed from the parcel. Terms in italics are second-stage additions to the core grouping. Roman numerals at right margin indicate Big Five factor that parcel is scored on in Modular Markers: I—Extraversion, II—Agreeableness, III—Conscientiousness, IV—Emotional Stability, V—Intellect. Minus sign after numeral indicates that parcel is reverse-scored.

parcels in Table 3 illustrates many of the kinds of attributes (e.g., Courage, Politeness, Predictability) that are not typically represented directly on Big Five measures.

Examination of the correlations of the 100 representative parcels with previous markers for the Big Five led me to conclude that creating orthogonal marker scales using these parcels alone would be difficult; there were too few parcels having sizable favorable direction loadings on one factor but

sizable unfavorable direction loadings on the other factor.[5] This indicates, of course, that frequently used personality descriptors tend to represent either favorable attributes or unfavorable ones, not evaluatively ambiguous ones. But the underlying dimension of favorable versus unfavorable characterization, typically confounded to some degree with desirability response style, tends to cause correlations among content scales (Jackson, 1970). Jackson (1970) noted the psychometric value of including some items reflecting minimal desirability variance in personality measures but noted that these are "rarely present in sufficient number, even in the most carefully written set of items" (p. 76). Some evaluatively ambiguous parcels (i.e., those with a combination of favorable and unfavorable loadings) would be needed to generate orthogonal marker scales. Because I was restricting myself to lexical stimuli, it was necessary to search further through available item pools rather than actually writing new items.

To create supplementary parcels, I made use of the full set of 587 adjectives originally administered to Samples A and C. (Most, but not all, of these adjectives were also administered to Samples B and D.) From this set, groups of terms were selected that (a) were not already in a representative parcel, (b) had some combination of favorable and unfavorable loadings on Big Five factors in previous studies, and (c) had relatively high correlations with one another. In general, these supplementary parcels are psychometrically less optimal than the representative parcels; they tend to be less homogeneous and therefore have either more items or lower reliability. Moreover, they draw on terms that are less frequently used and familiar to most respondents. However, they proved indispensable in building marker scales having mutual orthogonality. The 17 supplementary parcels eventually employed in the marker set are listed at the end of Table 3.

Selecting Parcels for the Marker Set

I used four criteria in selecting which parcels should be included in the marker set: (a) contribution to equidiscrimination for the overall scale, (b) maximum correlation with the criterion factor, (c) contribution to mutual

[5] Cattell and Tsujioka (1964) advocated development of "buffered tests employing suppressor action" (p. 5), on the principle that "a test, like a watch, shall have parts with different functions, properly balanced in certain ways" (p. 5). In plain terms, these authors advocated that markers for a factor should include items balanced with respect to their loadings on other factors. If one uses the structure matrix (correlations with factors rather than the oblique rotation pattern matrix that Cattell and Tsujioka suggested) to do the balancing, this approach tends to eventuate in mutually orthogonal scales. Guilford and Michael (1948) advocated attainment of univocal factor scoring by adding to high-loading variables one or more additional variables selected to suppress unwanted covariance with other factors, which is conceptually akin to Cattell's "buffering" but employed under the assumption that an orthogonal set of reference variables is the desired outcome.

orthogonality within the set of scales, and (d) brevity (using as few parcels as possible). The "criterion factor" was defined as the factor scores for the five-factor solution that best generalized across 14 data sets in the studies of Saucier (2000)—a Big Five structure based on several variable selections of personality adjectives as employed in self- and peer ratings of more than 3000 individual targets of description.[6] Factor scores for this five-factor solution were available in all of the samples used in the current study (A–F).

An "equidiscriminating test" (Nunnally, 1967; Nunnally & Bernstein, 1994) is able to measure well at all levels of the underlying dimension. Such a test could result simply from ensuring a wide diversity of item mean levels in a scale, but Nunnally recommended directly examining discrimination at differing levels of the dimension. In his approach, one selects items at multiple cutoff levels of the distribution; that is, items that are maximally discriminating at each of several points along the score continuum are selected and then combined to form an overall measure. This strategy tends to yield a spread of difficulties in the selected items (here, parcels), counteracting the tendency toward parallelism engendered by the use of reliability-maximizing and maximal factor-loading criteria (Saucier & Goldberg, in press a).

Criterion (a)—contribution to equidiscrimination—was applied by dichotomizing the distribution for each factor at each of five cut points: the 16.67% cumulative frequency point as well as the 33.33%, 50%, 66.67%, and 83.33% points. All of the parcels were correlated with each of the dichotomies, and the top three correlating parcels for each dichotomy were retained as an initial draft for the marker scale.[7] For example, for the Emotional Instability (i.e., Emotional Stability reversed) factor, Jealousy/Envy best discriminated at the high end, Anxiety did so at the middle cut point, and Emotional Excitability showed particularly good discrimination at the low end. Moodiness was also included in the initial draft; Jealousy/Envy, Anxiety, Emotional Excitability, and Moodiness were the only parcels to appear in the top three for the various dichotomies. The initial draft parcels for each factor were examined with respect to their mutual intercorrelations and their correlation with the criterion factor. Other parcels were examined with respect to whether they (a) might help to reduce scale intercorrelations and/or (b) were highly correlated with the residual of a linear regression using the initial

[6] These were "grand factor" scores based on secondary factor analyses of the factor scores derived from each of the 14 data sets, after all had been reproduced by regression equations in the largest sample in the study (see Saucier, 2000). The data sets included those with unipolar descriptors in both raw score and ipsatized form (representative selections of approximately 400 common personality adjectives) as well as those using bipolar scales (Peabody's [1987] representative set).

[7] The 33.33% and 66.67% cut points provided little incremental contribution to parcel selection beyond what the other three cut points provided; for the sake of parsimony, they were disregarded in this study.

draft marker scale to predict the criterion factor. I applied these criteria iteratively to successive drafts of each marker scale until I arrived at a set of marker scales that was difficult to improve on. For example, due to these considerations, the Moodiness parcel was eventually removed in favor of the Fretfulness and Hyperdevotedness parcels, resulting in a total of five parcels for the factor.

Big Five Modular Markers

The Big Three—a three-factor structure consisting of broad versions of Extroversion, Agreeableness, and Conscientiousness—has proved somewhat more robust than the Big Five in lexical studies of factor structures indigenous to multiple languages (Di Blas & Forzi, 1999; Saucier & Goldberg, in press b). I used the Big Three factors as the prime criterion and sought to build marker scales for the remaining two Big Five factors (Emotional Stability and Intellect) that were orthogonal to the Big Three.

The end result was a set of 7 parcels (20 items) for Extroversion, 9 parcels (22 items) for Agreeableness, and 7 parcels (18 items) for Conscientiousness. The other two Big Five factors were somewhat smaller, with 5 parcels (16 items) for Emotional Stability and 4 parcels (14 items) for Intellect. To help maximize the orthogonality of the Emotional Stability factor with respect to the other factors, I found it useful to include a parcel that consisted of amplifications (e.g., Overprotective, Overloyal) of common adjective descriptors, amplification terms that had been included in Goldberg's variable selections.[8] Items included in the 32 parcels can be discerned from an examination of the indicated Big Five parcels in Table 3.

The factor structure of these 32 parcels, which are based on 90 items, is illustrated in Table 4. It presents loadings for the 32 parcels in an exploratory principal factors analysis with communalities estimated via squared multiple correlations and varimax rotation (using Kaiser normalization) in a combined sample consisting of Samples A through E. The parcels reproduced the intended factors quite faithfully with quartimax as well as varimax rotation. In the promax solution (with kappa set to 4), all of the parcels had their highest pattern matrix loading on the intended factor, and intercorrelations among factors (when all were oriented by their more desirable poles) ranged from −.24 to .30, with a mean of .09 (mean of absolute values .15); thus, allowing the factors to be correlated had little effect on the obtained structure.

[8] Amplifications of desirable attributes can be helpful in creating orthogonal marker scales because they represent a mix of desirable and undesirable features. For example, the term *overloyal* suggests the desirable element of loyalty but the undesirable element of maladaptive excess with respect to the target or the degree of one's loyalties. The concept of attribute amplification was a crucial and fascinating part of Aristotle's conceptions of personality traits as discussed in his *Nicomachean Ethics.*

TABLE 4
Factor Structure of the 32 Parcels in the Big Five Modular Marker Set

	II	III	I	IV	V
Kindness	.77*	.26	.03	.00	−.22
Warmth	.73*	.20	.30	.08	.17
Sympathy	.73*	.21	.11	.18	.30
Agreeableness	.65*	.08	−.08	−.06	.26
Sensitivity	.60*	.16	.09	.43	.26
Toughness	−.64*	.04	.09	.10	.06
Slyness	−.55*	−.09	.13	.09	.16
Criticalness	−.47*	.05	.05	.25	.17
Demandingness	−.47*	.19	.22	.31	.11
Efficiency	.23	.79*	.03	−.11	.10
Organization	.07	.77*	−.04	−.05	−.01
Perfectionism	−.04	.71*	−.07	.12	.19
Decisiveness	.00	.55*	.21	−.31	.15
Caution	.21	.50*	−.31	.19	.10
Ambition	.08	.39*	.32	.03	.18
Forgetfulness	−.02	−.57*	−.01	.26	.01
Talkativeness	−.06	−.08	.70*	.11	−.13
Sociability	.35	.20	.66*	−.05	−.07
Assertiveness	−.37	.27	.62*	−.03	.20
Spontaneity	.18	−.16	.51*	.26	.31
Adventurousness	−.03	−.05	.47*	−.03	.34
Restraint	.10	.15	−.71*	.10	.07
Shyness	.11	−.08	−.66*	.23	.11
Fretfulness	−.16	−.20	−.22	.65*	−.08
Anxiety	−.20	−.12	−.01	.63*	−.02
Emotional Excitability	.18	−.06	.39	.59*	.08
Jealousy/Envy	−.34	−.20	.02	.55*	−.11
Hyperdevotedness	.13	.12	−.14	.48*	.09
Analytical Inquiry	.01	.15	−.02	.05	.81*
Reflectiveness	.20	.11	−.16	.07	.65*
Intellectuality	.11	.32	.09	−.12	.52*
Unconventionality	−.24	−.41	.21	−.03	.41*

Note. $N = 1620$. Coefficients are varimax-rotated factor loadings. I—Extraversion (Dynamism), II—Agreeableness (Altruism vs Antagonism), III—Conscientiousness (Self-Regulation), IV—Emotional Stability (reversed, Anxiety), V—(Autonomous) Intellect.

* Highest loading for variable.

Obviously, the parcels included in the Modular Markers for the Big Five were selected by a more elaborate procedure than that used for the Ortho-40 approximation described earlier. This second set was derived in large part from a representative selection of variables and in a process that involved equidiscriminating criteria. The delineation of subcomponent parcels is itself an advantage; these parcels enable one to identify more specific sources of variation and cancel out some of the artifactual variance associated with aggregating or factor-analyzing single items. Are the Modular Markers an improvement on the earlier approximation in other ways?

Reliability and Interscale Correlations

Table 5 presents coefficient alpha internal consistency estimates for the Modular Markers, which are higher (by an average of about .05) than those for the Ortho-40. Table 5 also presents interscale correlations for the Big Five Modular Markers, including comparisons to interscale correlations in

TABLE 5
Reliability and Interscale Correlations of Modular Marker Scales

	Derivation samples			Cross-validation sample
	Self	Liked Peer	Pooled Peer	Community Sample
Coefficient alpha				
I	.88	.89	.91	.84
II	.82	.86	.94	.83
III	.85	.88	.91	.86
IV	.79	.75	.80	.82
V	.77	.75	.87	.82
Interscale correlations				
I–II	−.11	−.07	−.02	−.01
I–III	.11	.04	−.03	.22
I–IV	−.06	−.05	−.10	.08
I–V	.13	.26	.15	.20
II–III	.01	.12	−.01	.05
II–IV	−.09	.04	.28	.28
II–V	.02	.00	.21	−.21
III–IV	.01	.13	.09	.23
III–V	.04	.02	.23	.03
IV–V	.00	.00	.18	−.02
Modular markers mean	.01	.05	.10	.08
100 Markers mean	.13	.24	.27	.25
Mini-Markers mean	.11	.18	.26	.22
Ortho-40 mean	.01	.05	.09	.07

Note. Sample sizes: Self = 320, Liked Peer = 316, Pooled Peer = 205, Community Sample = 592. All analyses used the original (nonipsatized) response data.

the same samples for Goldberg's 100 Markers.[9] Goldberg's marker scales demonstrate the typical level of scale intercorrelation for previous Big Five marker sets, with an overall average (across samples and pairs of factors) of .21. By contrast, the Big Five Modular Markers have an overall average scale intercorrelation of only .05, and for only one pair of factors (Extroversion–Intellect, .18) does the mean intercorrelation across five samples exceed .10. Interscale correlations are about as low as those found for the Ortho-40, despite higher reliability.

Thus, the Modular Markers provide an even more powerful demonstration that virtually orthogonal Big Five marker scales are a realizable goal. This tends to validate one core assumption of the Big Five model: The independence of the factors, which results from the constraints of factor-analytic algorithms, can be practically operationalized in a measurement instrument that uses simple straightforward scoring (rather than ipsatization of subject responses or elaborate scale scoring based on factor score weights).

A SHORT FORM OF THE BIG FIVE MODULAR MARKERS

Many useful criteria might conceivably be applied to item selection in the scale construction process, and some of these criteria tend to conflict with others. For example, reliability and brevity criteria tend to lead one to long and short scales, respectively, and mutual orthogonality of personality factor marker scales usually demands that one depart from a purely representative sampling of variables (Saucier & Goldberg, in press a). Because of such trade-offs, there can be no single perfect marker set for a factor structure. A disadvantage of the Big Five Modular Markers is length; a total of 90 items are required. Can this marker set be abbreviated to 40 items while maintaining low interscale correlations?

Table 6 provides reliability estimates and interscale correlations for a set of 40 Mini-Modular Markers (the 3M40). This reduced set of adjectives was developed by selecting from the 90 Big Five Modular Marker terms a subset of terms that (a) retained the highest loading items with (b) about equal numbers having positive and negative loadings on each of the other factors, (c) while maintaining a spread of response means on each scale, with some secondary attention also to (d) maintaining balanced keying, (e) representing as many of the 32 parcels as is feasible, and (f) excluding items where doing so increased the reliability of the scale. Of the 40 3M40 items, 22 are found also on the Ortho-40 and thus on all three of the marker sets presented in this article.[10]

[9] The sample size for Sample E was reduced to 592, the number who used all 791 adjectives administered at the three times (1993, 1995, and 1998).

[10] The 22 terms found in all three marker sets were as follows: for Extroversion, Talkative versus Quiet, Shy, and Reserved; for Agreeableness, Kind and Sympathetic versus Cold, Harsh, and Demanding; for Conscientiousness, Efficient and Organized versus Disorganized;

TABLE 6
Mini-Modular Markers (3M40): Reliability and Interscale Correlations

	Derivation samples			Cross-validation sample
	Self	Liked Peer	Pooled Peer	Community Sample
Coefficient alpha				
I	.82	.84	.85	.77
II	.71	.76	.89	.71
III	.76	.75	.84	.76
IV	.67	.63	.71	.72
V	.67	.64	.80	.73
Interscale correlations				
I–II	.02	.05	.01	.09
I–III	.03	−.04	−.06	.19
I–IV	−.05	−.09	−.17	.06
I–V	.09	.15	.07	.14
II–III	.01	.08	−.04	.10
II–IV	−.05	.04	.26	.24
II–V	.00	.10	.24	−.10
III–IV	.03	.09	.06	.18
III–V	−.02	−.02	.08	−.04
IV–V	−.02	.06	.17	.08
3M40 mean	.01	.04	.06	.10
Ortho-40 mean	.01	.05	.09	.07
100 Markers mean	.13	.24	.27	.25
Mini-Markers mean	.11	.18	.26	.22

Note. Self sample N = 320, Liked Peer sample N = 316, Pooled Peer sample N = 205, Community Sample, N = 592. All analyses were conducted with raw (nonipsatized) data. Scale items are as follows: I—Assertive, Playful, Sociable, Talkative versus Quiet, Reserved, Shy, Withdrawn; II—Kind, Sentimental, Sympathetic, Tolerant versus Cold, Critical, Demanding, Harsh; III—Cautious, Efficient, Meticulous, Organized, Perfectionistic versus Absent-minded, Disorganized, Indecisive; IV—Unenvious, Unexcitable versus Anxious, Emotional, Fearful, Fretful, High-strung, Nervous; V—Complex, Intellectual, Nonconforming, Philosophical, Unconventional versus Conventional, Unintellectual, Unreflective.

Compared to the full set of 90 Modular Marker items, interscale correlations for the 3M40 are about the same on average.[11] But reliability is lower (almost .10 per scale on average) than for the longer marker set. Compared to the Ortho-40 described earlier, interscale correlations are similar, but reliability for the 3M40 scales is slightly lower (generally by less than .05).

for Emotional Stability, Unenvious and Unexcitable versus Anxious, Fretful, Fearful, Nervous, and Emotional; for Intellect, Complex and Philosophical versus Unreflective.

[11] When one calculates, instead of the means of the obtained values, the means of the absolute values of scale intercorrelations, the figures are .03, .07, .12, and .12 for the 3M40 in the Self, Liked Peer, Pooled Peer, and Community samples, respectively, as compared to .06, .07, .13, and .13 for the Modular Markers.

The lower internal consistency is probably related to the higher degree of representative sampling in the 3M40 scales. Although the two marker sets have nearly identical items for Factor IV, on the other factors, the 3M40 scales appear to be broader in content reference, primarily because the item pool in the Modular Markers preserves a broader range of content than is found in the 100 Markers on which the Ortho-40 is based. For example, the 3M40's scale for Intellect has "unconventionality" content that is lacking in the Ortho-40 version (as well as in the 100 Markers). Representative sampling does not maximize coefficient alpha, although it may heighten validity with respect to a broad array of criteria. Indeed, validity is the final issue I must examine with respect to these new, more mutually orthogonal Big Five marker scales.

ARE MUTUALLY ORTHOGONAL MARKER SETS AS VALID AS PREVIOUS MARKER SETS?

Validity ultimately pertains to the demonstrated value of an instrument for a given purpose; such value can be demonstrated in a diversity of ways, which is one reason why "construct validity" is unlikely to be established by a single study and instead requires a long, perhaps unending, validation process.

One element of adequate validity is a congruence between what the instrument is *believed* (or theorized) to measure and what the instrument is *demonstrated* to measure. I noted earlier that the Big Five factors have been *theorized* to be independent factors and that the failure of marker sets to *demonstrate* mutually orthogonality among the factors creates a validity issue jointly for these marker sets and for the theoretical standpoints they are supposed to operationalize. For one who makes the theoretical assumption that the Big Five are mutually orthogonal, the marker sets presented in this article do not generate these particular question marks with respect to validity. Therefore, the Ortho-40, Modular Markers, and 3M40 might be judged superior to previous Big Five measures in one aspect of validity.

But there are other aspects of validity. In comparison to previous marker sets, do these new marker sets "measure up" in terms of concurrent and predictive validity? Or are there indications that mutual orthogonality compromises validity?

Correlations with Criterion Lexical Factors

How well do the marker scales reflect factors found in analyses of large numbers of personality adjectives in English? Saucier (2000) has provided the most canonical (based on the largest collections of data) version of the Big Five based on English-language adjectives. Sample E, a sample not used for final selection of any of the items on any of the new marker sets, is an appropriate sample in which to make comparisons; a total of 431 participants

in this sample provided data that allowed scoring for the Modular Markers, the 100 Markers and Ortho-40, and the NEO-PI-R (Costa & McCrae, 1992), the most commonly used of the five-factor measures, as well as the NEO–FFI (a short form of the NEO-PI-R) and the Big Five Inventory (Benet-Martínez & John, 1998). I correlated scale scores from these five-factor marker sets with factor scores for the lexical Big Five derived from Saucier (2000).

In descending order of average magnitude, the correlations between matched pairs consisting of a Big Five factor and a corresponding Big Five scale ranged from .77 to .89 (mean .83) for the Modular Markers, from .75 to .86 (mean .80) for the 3M40, from .67 to .81 (mean .70) for the 100 Markers, from .65 to .75 (mean .69) for the Ortho-40, from .55 to .76 (mean .67) for the Mini-Markers (a 40-item short form of the 100 Markers [Saucier, 1994a]), from .56 to .76 (mean .67) for the NEO-PI-R, from .60 to .72 (mean .66) for the NEO-FFI, and from .56 to .75 (mean .65) for the BFI. These correlations all are high but tend to favor the Modular Markers and 3M40; this would be expected because these marker sets were designed based on the version of the Big Five found by Saucier (2000), the same version of the Big Five represented in this criterion. Moreover, this is not an adequate validity comparison because the same adjectives used to compute scale scores were among those used to derive Saucier's factors. So I turned instead to other criteria based on additional measures administered to 394 members of Sample E.

Correlation with Self-Reported Frequencies of Behavioral Acts

One useful validity criterion would involve judgments as to the frequencies of concrete behaviors rather than judgments with respect to global attributes. Goldberg (in press) employed such criteria in studies comparing the validity of a range of personality inventories. The criteria were based on self-reported frequencies of behavioral acts. Participants in the current Sample E were sent a list of 400 behavioral acts (e.g., played chess, shot a gun, polished my toenails); most of these were from a classic study of Loehlin and Nichols (1976), with the remainder added by Goldberg and his research team. For each act, participants reported the frequency with which they had carried out that activity, using the following response options: (1) never in my life, (2) not in the past year, (3) once or twice in the past year, (4) three or more times in the past year but not more than 15 times, or (5) more than 15 times in the past year. Goldberg cluster-analyzed the 400 activities by various methods, leading to multi-act clusters of related behaviors.

I employed the same six behavior clusters that Goldberg (in press) focused on. The clusters were Creative Achievements (11 acts, e.g., acted in a play, produced a work of art, played a piano or other instrument), Friendship (8 acts, e.g., hugged someone, complimented someone, did a favor for a friend),

Writing (8 acts, e.g., wrote a thank you note, made an entry in a diary or journal, worked on a scrapbook), Reading (6 acts, e.g., bought a book, read an entire book in one sitting, went to a public library), Undependability (7 acts, e.g., broke a promise, let work pile up until just before a deadline, arrived at an event more than an hour late), and Drug Use (14 acts, e.g., smoked marijuana, had a hangover, drank alcohol during working hours). Coefficient alpha reliabilities for these clusters ranged from .70 to .89. Two of the behavior clusters involved relatively undesirable activities (Drug Use and Undependability), two involved relatively desirable activities (Friendship and Creative Achievements), and two involved more evaluatively ambiguous activities (Reading and Writing).

Which of the Big Five measures should best predict these criteria? Based on the common assumption that higher reliability enables higher validity, one would predict that longer scales having higher coefficient alpha values would have higher predictive validity coefficients. On this basis, the NEO-PI-R domain scale scores (based on 48 items per scale) should have the highest validity, with the 100 Markers (with 20 items per scale) ranking next. The most crucial comparison for our purposes is between the mutually orthogonal marker sets of scales (i.e., Ortho-40, Modular Markers, and 3M40) and the other marker sets that have higher interscale correlations.

The top half of Table 7 shows multiple correlations (values of *R*) between each of these criteria and all five scales in each of eight Big Five marker sets. The NEO-PI-R, as predicted, had the highest average validity coefficient (.41), but it was not ahead by much, and for three of the six criteria at least one other marker set had a higher multiple correlation. There was little difference between validity coefficients for the other marker sets, which ranged from .36 to .39 on average. Average validity coefficients for the three mutually orthogonal marker sets were as high as (in fact, very slightly higher than) those for the BFI, 100 Markers, and Mini-Markers. These findings provide one indication that a mutually orthogonal marker set does not compromise validity.

Correlation with Peer Ratings on Another Big Five Measure Administered Separately

One limitation of the behavior frequency criteria is that, like the predictor scores, they are based entirely on self-ratings; it is possible that the same self-perception biases are affecting both predictor and criterion. Thus, criteria based on peer ratings would be a useful complement. Members of Sample E had also been described by up to three acquaintances nominated by themselves, using a peer-rating version of the Big Five Inventory. Because BFI scale scores based on these peer ratings are substantially (more than .20 on average) intercorrelated, I performed a principal factors analysis of the ratings (original, not ipsatized) provided by a total of 1476 peers of sam-

TABLE 7
Multiple Correlations between Self-Ratings from Five-Factor Marker Sets and Behavior-Frequency and Peer-Rating Criteria

	Questionnaires			From pool of 100 marker adjectives			From pool of modular marker adjectives		
Criterion	NEO-PI-R	NEO-FFI	BFI	100 Markers	Mini-Markers	Ortho-40	Modular Markers	3M40	Mean
Behavior clusters									
Drug Use	.41*	.35	.29	.31	.29	.31	.33	.31	.32
Friendship	.44*	.42	.40	.36	.39	.41	.37	.40	.40
Undependability	.32	.34	.27	.30	.29	.30	.37	.38*	.32
Reading	.37	.36	.32	.36	.36	.37	.41*	.39	.37
Writing	.44*	.42	.34	.34	.37	.39	.36	.36	.38
Creative Achievements	.45	.43	.47*	.46	.44	.46	.41	.41	.42
Mean	.41*	.39	.35	.36	.36	.37	.37	.37	
BFI peer factors									
Extroversion	.56	.52		.66*	.64	.65	.64	.65	.62
Agreeableness	.47	.46		.41	.43	.38	.49*	.42	.44
Conscientiousness	.51	.50		.50	.50	.49	.51*	.50	.50
Emotional Stability	.47*	.44		.42	.34	.43	.38	.41	.41
Openness/Intellect	.52	.51		.56*	.55	.56*	.51	.50	.53
Mean	.51	.48		.51*	.49	.50	.50	.50	

Note. $N = 394$, Sample E. 3M40, Mini-Modular Markers. Behavior clusters are clusters of behavioral acts used by Goldberg (in press). BFI peer factors are derived from varimax-rotated common factor analysis of 44 Big Five Inventory items in a set of ratings by 1476 peers nominated by the targets in Sample E, with factor scores averaged to create one set of factor scores per target.

* Highest coefficient for each criterion (row).

ple members and retained the regression-based factor scores based on five varimax-rotated factors. Sets of factor scores estimated for the same target of description were averaged; I used these averaged factor scores for each of the five factors as criteria. The BFI self-ratings were not used as predictors because the method variance they share with BFI peer ratings would make comparisons with other instruments difficult.

The bottom half of Table 7 shows the multiple correlations for the entire set of five predictors in each marker set, with each peer-rating factor as criterion. One would again expect the long-form NEO-PI-R to have the highest coefficients, but its 48-item scales had little or no advantage over the adjective-based marker sets, which in turn all had an average validity coefficient of about .50. Once again, the validity coefficients for the mutually orthogonal marker sets were not lower than those with more intercorrelated scales, even though, as described earlier, the sets of scales with higher scale intercorrelations tend to have higher alpha coefficients.

Discussion of Validity Analyses

Reliability is customarily viewed as a prerequisite to validity. It is commonly assumed that by making one's measure more homogeneous (i.e., internally consistent), one is indirectly improving its validity. But these results illustrate that, in the item selection process, one does not always attenuate validity when one attenuates reliability. Holding scale length equal, the scales in mutually orthogonal marker sets tend to have lower reliability than their more intercorrelated counterparts, but their validity does not seem to be attenuated in a commensurate fashion. Why? Perhaps this occurs because a set of mutually orthogonal scales minimizes multicollinearity and thus may preserve range and breadth of content better than do sets of factor scales that are allowed to be intercorrelated. It may be that attempts to make scales mutually orthogonal require one to diversify the content representation of the scales to a greater degree than if one were simply maximizing reliability or factorial replicability. In other words, mutually orthogonal scales may be less prone to be narrow in content and therefore less prone to generate the classic attenuation paradox (Loevinger, 1954).

A striking feature of Table 7 is the relative homogeneity of the validity coefficients when comparing one Big Five measure with another. From a validity standpoint, it seems to make little difference which of these marker sets one uses. Increasing the number of items by a factor of 4 (as in going from the 60-item NEO-FFI to the 240-item NEO-PI-R) seems to do relatively little to increase validity. And making a set of scales orthogonal might not lessen validity even in cases when it lessens coefficient alpha reliability.

CONCLUSIONS

Compared to previous markers and measures of the Big Five, these new marker sets have much lower interscale correlations. One of them (the Modu-

lar Markers) includes an explicit set of subcomponent parcels, whereas the other two (the Ortho-40 and 3M40) provide a relatively brief assessment, requiring but 40 items instead of 90. The internal consistency (coefficient alpha) of these mutually orthogonal marker sets is slightly lower than is characteristic of Big Five marker sets that have higher scale intercorrelations. However, in this case, lower reliability does not seem to translate into lower validity; the Modular Markers, Ortho-40, and 3M40 showed predictive validities comparable to those of comparison Big Five marker sets. I have suggested that mutually orthogonal scales minimize multicollinearity and thus preserve range and breadth of content representation to a higher degree than do sets of scales with more multicollinearity.

A major discrepancy between theory and practice with respect to the Big Five personality factors has been the use of marker sets for the factors that have substantial scale intercorrelations, although the factors are posited to be independent. These studies indicate that the nonorthogonality of previous Big Five measures is an accident of scale construction rather than a property of the factors. One can create mutually orthogonal scales for the Big Five if one sets a priority for doing so and goes beyond the most commonly used reliability-maximizing approaches to scale construction. The Big Five as measured by these mutually orthogonal marker sets do not differ substantially from previous conceptualizations of the Big Five. The scales tend to be less reliable than, but appear to be just as valid as, previous measures of comparable scale length.

Accordingly, Block's (1995) critique of the nonorthogonality of Big Five measures can now be focused more directly on the real source of nonorthogonality—not the Big Five factors themselves, but rather the scale construction procedures used by developers of the measures. These procedures did not go as far as they might have to ensure that orthogonal factors (in theory) have orthogonal markers (in practice). The current studies eventuated in three different adjectival marker sets for the Big Five; one used the item pool of a larger and frequently used Big Five marker set, another was based on parcels built from a representative sampling of the personality domain, and a third was a short form of the parcel-based marker set. Comparable results from the three marker sets indicate that there is more than one way to attain the goal of mutually orthogonal scales.

The search for a scientifically compelling taxonomy of personality attributes currently involves important controversies about which structure of broad factors is most optimal. The Big Five is the most salient candidate model, but in science the most salient model in a field should be subjected to test, trial, and criticism because this leads to the best eventual outcome—either validation of the model or replacement of the model by one more suitable. The Big Five may eventually fall by the wayside and be replaced, but the reason for doing so should not be that Big Five measures have in-

cluded nonorthogonal sets of scales. Nonorthogonal Big Five scales can now be seen as a potentially superfluous feature of particular Big Five measures rather than a fundamental problem with the constructs in the Big Five model.

REFERENCES

Benet-Martínez, V., & John, O. P. (1998). Los Cincos Grandes across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology,* **75,** 729–750.

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

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* Yonkers-on-Hudson, NY: World Book.

Cattell, R. B., & Tsujioka, B. (1964). The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales. *Educational and Psychological Measurement,* **24,** 3–30.

Clemans, W. V. (1966). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs,* **14.**

Costa, P. T., & 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.

Di Blas, L., & Forzi, M. (1999). Refining an Italian trait taxonomy by integrating simple structure and circumplex models: The abridged Big Three circumplex structure. *Journal of Personality and Social Psychology,* **76,** 451–481.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 41, pp. 417–440). Palo Alto, CA: Annual Reviews.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology,* **73,** 1246–1256.

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

Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs,* **7,** No. 72-2.

Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 1, pp. 203–234).

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology,* **59,** 1216–1229.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment,* **4,** 26–42.

Goldberg, L. R. (in press). The comparative validity of adult personality inventories: Applications of a consumer-testing framework. In S. R. Briggs, J. M. Cheek, & E. M. Donahue (Eds.), *Handbook of adult personality inventories.* New York: Plenum.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw–Hill.

Guilford, J. P., & Michael, W. B. (1948). Approaches to univocal factor scores. *Psychometrika,* **13,** 1–22.

Hofstee, W. K. B. (1998). How to score questionnaires. *Personality and Individual Differences,* **25,** 897–909.

Hofstee, W. K. B., De Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology,* **63,** 146–163.

Jackson, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 61–96). New York: Academic Press.

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review,* **78,** 229–248.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: Guilford.

Loevinger, G. (1954). The attenuation paradox in test theory. *Psychological Bulletin,* **51,** 493–504.

McCrae, R. R., Zonderman, A. B., Costa, P. T., 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.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw–Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw–Hill.

Peabody, D. (1987). Selecting representative trait adjectives. *Journal of Personality and Social Psychology,* **52,** 59–71.

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology,* **57,** 552–567.

Rogler, L. H. (1999). Methodological sources of cultural insensitivity in mental health research. *American Psychologist,* **54,** 424–433.

Saucier, G. (1992). Benchmarks: Integrating affective and interpersonal circles with the Big-Five personality factors. *Journal of Personality and Social Psychology,* **62,** 1025–1035.

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

Saucier, G. (1994b). Separating description and evaluation in the structure of personality attributes. *Journal of Personality and Social Psychology,* **66,** 141–154.

Saucier, G. (1997). Effects of variable selection on the factor structure of person-descriptors. *Journal of Personality and Social Psychology,* **73,** 1296–1312.

Saucier, G. (2000). *A hierarchy of personality-attribute dimensions: Integrating the Big Five and broader-factor structures.* Submitted.

Saucier, G., & Goldberg, L. R. (1998). What is beyond the Big Five? *Journal of Personality,* **66,** 495–524.

Saucier, G., & Goldberg, L. R. (in press a). 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.* Goettingen, Germany: Hogrefe & Huber.

Saucier, G., & Goldberg, L. R. (in press b). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality.*

Saucier, G., Ostendorf, F., & Peabody, D. (2001). The non-evaluative circumplex of personality adjectives. *Journal of Personality,* **69,** 537–582.

Schmitt, D. P., & Buss, D. M. (2000). Sexual dimensions of person description: Beyond or subsumed by the Big Five? *Journal of Research in Personality,* **34,** 141–177.

ten Berge, J. M. F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. *Multivariate Behavioral Research,* **34,** 89–102.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology,* **37,** 395–412.