The Comparative Validity of Questionnaire Data (16PF Scales) and Objective Test Data (O-A Battery) in Predicting Five Peer-Rating Criteria

Lewis R. Goldberg
University of Oregon and Institute for the Measurement of Personality
Warren T. Norman and Edward Schwartz
University of Michigan

Thirty tests from the 1955 edition of Cattell's Objective-Analytic (O-A) Test Battery, plus Forms A and B of the Sixteen Personality Factor Questionnaire (16PF), were administered to 82 male undergraduates. In addition, each subject was rated by 7 to 11 close associates on each of 20 bipolar rating scales, 4 scales tapping each of 5 peer-rating factors. These peer ratings were used as criterion variables to be predicted by the 16PF scales and by the O-A Battery. The O-A Battery measures were slightly more highly related to one peer-rating factor (Culture); the 16PF scales were slightly more highly related to another (Conscientiousness); and the two sets of test variables were essentially equivalent in predicting the other three factors (two of which showed no significant relationships with either instrument). The lack of any consistent superiority of the objective test scores over the questionnaire scales, coupled with some criticisms of the objective tests on purely logical grounds, should make one cautious in accepting the claims being made for the comparative validity of the O-A Battery.

Individual differences can be assessed in at least three different media, namely via Life (L) data, Questionnaire (Q) data, or Objective Test (T) data (see Cattell, 1957). The distinction between L data and Q or T data is reasonably clear: L data refers to behavior in situ (within the ongoing stream of life), whereas both Q and T data are constrained by experimental stimulation or laboratory control. Within the traditional psychometric paradigm, L data are often assigned the critical role of the dependent variables or the criterion behaviors to be predicted, and these predictions are typically generated from measures based on either Q or T data.

While the field of psychological assessment has become heavily saturated with personality questionnaires and inventories (see Goldberg, 1971), Cattell has consistently exhorted psychometricians to eschew Q data and to concentrate on the development of objective personality tests. To this end, Cattell and his associates have constructed a multitude of such tests over the years and have included these measures, plus many devised by other investigators, in a series of factorial studies (see Cattell, 1957; Cattell & Warburton, 1967; Hundleby, Pawlik, & Cattell, 1965). The explicit rationale for this research strategy is the assumption that objective tests eliminate many of the biases that presumably distort questionnaire responses (e.g., response sets and styles) and, consequently, that such tests will ultimately provide the most valid measures of important personality traits.

By 1955 Cattell had assembled a battery of objective tests called the Objective-Analytic (O-A) Eighteen Factor Personality Test Battery, which was distributed by the Institute for Personality and Ability Testing (IPAT) as a preliminary re-

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 4, No. 2 Spring 1980 pp. 183–194

search kit. This initial battery was probably not widely disseminated; perhaps because of its limited availability and the relatively high cost of using objective tests as compared to questionnaires, there have been very few reports of studies by investigators outside Cattell's laboratory comparing the validities of these objective tests with those obtained from questionnaire scales. The present report summarizes the results of one such comparison, using a version of the O-A Battery that included 30 tests and yielded 57 scores. At the present time the number of tests catalogued by Cattell and Warburton (1967) numbers over 400, yielding well over 2,000 scores. Consequently, the results presented in this report must be understood as limited to only a subset of the total set of objective tests currently available.

Procedure

Overview

As part of a larger study (Norman, 1963), 82 male undergraduate students from 9 fraternity houses at the University of Michigan were administered 30 tests from the 1955 edition of the O-A Battery, plus Forms A and B of the 16 Personality Factor Questionnaire (16PF). In addition, each student was rated by 7 to 11 close associates on each of 20 bipolar rating scales, 4 scales tapping each of the 5 major peer-rating factors found in previous L-data analyses (Norman, 1963; Tupes & Christal, 1961). These peer ratings were averaged across raters and across the 4 defining scales to provide 5 peer-rating composite scores, which were used as the criterion variables in the present study. The validity of predictions of these criteria from the 16PF scales (Q data) was compared with that achieved from the O-A Battery (T data).

Subjects

The students were primarily college seniors, all males, who lived in fraternity houses located near the campus of the University of Michigan. All students were paid for their participation in the study. Four testing sessions were conducted in the same order for all students. The first session was held in each of the fraternity houses, where participating members of the fraternity sat around a table in full view of one another and rated each other on the 20 peer-rating scales. Between the first and second testing sessions, all students completed Forms A and B of the 16PF plus other questionnaires not analyzed in the present study. Subsequently, each student was administered the O-A Battery during three testing sessions, each of 3 hours in duration, each session separated by 1 week.

The Five Peer-Rating Criteria

The 20 bipolar rating scales—4 marking each of 5 orthogonal rating factors—were developed initially from the comprehensive collation of trait names in the English language (Allport & Odbert, 1936), which was subsequently reduced, clustered, and factor analyzed by Cattell (1947, 1957) and more recently was studied intensively by other investigators, including Tupes and Christal (1961), Norman (1963), and Norman and Goldberg (1966). These five peer-rating factors, which served as criterion variables in the present project, have been named as follows: (1) Surgency or Extroversion, (2) Agreeableness, (3) Dependability or Conscientiousness, (4) Neuroticism versus Emotional Stability, and (5) Culture.

The peer ratings were carried out separately within each of 9 groups (one of 7 members, two of 8, two of 9, three of 10, and one of 11), the members of each group having lived together in close association for periods of at least 1 to 3 years. Within each rating group every student rated all of the others in his group on each of the 20 bipolar scales using a forced distribution; students selected a designated number of their group (approximately one third) as being most highly characterized by the positive pole of the scale, then chose the same number as being closest to the negative pole, and finally placed the remaining members in a middle category. All

Table 1
Intercorrelations Among the Five Peer-Rating Factors (N = 82)

	Peer-Rating Factor			
Factor	II	III	IV	V
I	.19	-.28*	-.03	.25*
II		.14	.46**	.10
III			-.06	.22*
IV				-.16

*p < .05; **p < .01

members of the group were rated on one scale before the rater proceeded to the next scale. The two poles of each scale were described on the rating forms, and the four scales marking a particular peer-rating factor were presented on every fifth page of the rating booklet.

The 16 PF

Both Forms A and B of the 16PF were completed by each of the 82 students, working alone in their fraternity houses, between the first and the second testing sessions. For a description of the scales included in this inventory, see Cattell, Eber, and Tatsuoka (1970).

The O-A Battery

From the 1955 version of the O-A battery, all tests that were already prepared, or could be adapted, for paper-and-pencil group testing were first administered to 24 paid college students (other than those included in the present study). The aims of this pretest were (1) to familiarize the test administrator with the battery, (2) to develop scoring stencils and keys, (3) to simplify and clarify the test instructions, (4) to estimate the time required for each test in the battery, and (5) to eliminate from the final test battery any of the tests either presenting unusual difficulties in administration and/or scoring or providing scores that did not differentiate among college students. On the basis of the pretest, 9 test variables were eliminated from the final battery—2 because of scoring ambiguity and 7 because of poor discrimination in the pretest sample. The remaining 57 O-A Battery measures were scored from 30 tests; of these, 16 provided only one score, 8 provided two scores, 4 provided three scores, and 2 provided four or more scores.

Results

Within-Media Analyses

Although the original peer-rating factors were orthogonal (Tupes & Christal, 1961), there are typically some moderate intercorrelations among the composite scores formed by averaging across the four marker scales for each factor (see Norman, 1963). The intercorrelations among these peer-rating composite scores in the present study are listed in Table 1. The highest of these correlations (r = .46) was between Factors II (Agreeableness) and IV (Emotional Stability). Three of the other nine correlations were significantly greater than zero ($p < .05$), but they were all of small size.

The parallel-form equivalence correlations for each of the 16PF scales are listed in Table 2, along with the estimated reliabilities of each of the 16PF composite scores when Forms A and B were combined.[1] The parallel form correlations, which averaged .37, ranged from a low of − .09 for Factor N (Artless versus Shrewd) to a high of

[1]The intercorrelations among the 16PF scales, computed separately within Forms A and B, are available from the first author, along with the between-form correlations.

Table 2
Correlations between 16PF Forms A and B, and Estimated Reliability of the Composite Scores (N = 82)

Scale	Correlation A vs. B	Reliability A + B*
A	.50	.67
B	.37	.54
C	.37	.54
E	.40	.57
F	.54	.70
G	.37	.54
H	.71	.83
I	.30	.46
L	.53	.69
M	.23	.37
N	-.09	--
O	.32	.48
Q1	.14	.25
Q2	.29	.45
Q3	.36	.53
Q4	.57	.73

*Estimated by the Spearman-Brown formula.

.71 for Factor H (Shy versus Venturesome). These equivalence coefficients are somewhat lower than, though highly correlated with, those reported in the 16PF Handbook (Cattell, Eber, & Tatsuoka, 1970: Table 5.3, p. 33). As has been pointed out repeatedly in reviews of the 16PF (e.g., Buros, 1970), a number of 16PF scales have consistently higher between-form correlations with other scales than with those to which they are supposedly equivalent.

The means, standard deviations, and intercorrelations among the composite 16PF scores (Forms A and B combined) are presented in Table 3. Almost 15% of these 16PF scale intercorrelations were greater than .40, and some of them were quite high (e.g., the correlation between Factors O and Q4 was .73, and that between Factors C and Q4 was −.68).

Of the 1,596 correlations among the 57 O-A Battery variables, less than 4% were greater than .30 and only 1% were above .40, most of the latter being instances of two variables scored from the same test. A number of factor analyses of these 57 variables, each employing a different criterion for estimating the number of factors or for rotating the resulting factors, produced quite similar results. When 18 factors were rotated to an oblique simple structure criterion, all of the intercorrelations among the oblique factors were of trivial magnitude, none being significantly greater than zero. Moreover, the few variables with high loadings on each factor were often those scored either from the same test or from two very similar tests; indeed, variables scored from the same test provided the highest loadings on 7 of the 18 factors.

Across-Media Analyses: 16PF versus O-A Battery

Of the 912 correlations between each of the 16PF scales and the 57 variables scored from the

Table 3

Means, Standard Deviations, and Intercorrelations among the 16PF Scales (Combined A and B Forms) (N = 82)

Scale	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
A		.03	-.05	-.01	.18	-.11	.26*	.18	.04	-.19	.12	.05	-.07	-.37**	-.03	.02
B			.03	.10	.18	-.20	.01	-.02	-.13	.12	.32**	.01	.13	.12	-.21	-.02
C				.10	.23*	-.05	.42**	-.18	-.53**	-.32**	.28**	-.64**	.07	-.11	.46**	-.68**
E					.51**	-.27*	.48**	.27*	-.07	.29**	.27*	-.22*	.10	.02	-.15	-.03
F						-.28**	.47**	.27*	-.05	.14	.11	-.18	.07	-.24*	-.09	-.08
G							-.11	-.36**	.26*	-.27*	-.01	.04	-.32**	-.11	.25*	.09
H								-.02	-.33**	-.14	.37**	-.59**	.07	-.43**	.20	-.39**
I									-.06	.46**	-.15	.23*	-.02	.00	-.19	.18
L										.27*	-.14	.48**	-.15	.03	-.35**	.55**
M											-.20	.36**	.25*	.33**	-.37**	.36**
N												-.38**	.20	-.03	.24*	-.37**
O													-.17	.19	-.55**	.73**
Q1														.32**	-.13	-.15
Q2															-.12	.18
Q3																-.55**
Mean	18.9	19.1	32.0	28.0	27.8	22.5	26.1	20.1	17.3	22.0	23.2	20.7	21.4	20.5	19.5	27.5
σ	6.7	2.8	6.0	5.8	7.5	5.0	8.9	5.4	5.9	5.6	3.9	6.4	4.3	5.1	5.4	7.4

*p < .05; **p < .01

O-A Battery[2], less than 4% were greater than .30, and only two (M. I. 246 from the O-A Battery with 16PF Factors G and Q1) were greater than .40. About 11% of these correlations were significantly greater than zero at $p < .05$, and about 5% were significant at $p < .01$. Of the 16PF scales, Factor M provided the highest number of significant correlations (13) with the O-A Battery variables, and Factor Q2 provided the least (2). From the O-A Battery, M. I. 246 provided the highest number of significant correlations (6), whereas 15 of the O-A Battery variables were not significantly correlated with any of the 16PF scales.

Across-Media Analyses: Prediction of the Peer-Rating Criteria

The significant 16PF and O-A Battery correlates[3] of the five peer-rating factors are summarized in Tables 4 and 5. Perhaps the most striking aspect of Table 4 is the differential predictability among the five peer-rating factors. Only a chance-level sprinkling of variables provided significant correlations with Factors II (Agreeableness) and IV (Neuroticism versus Emotional Stability). At the other extreme, there were a considerable number of significant predictors of Factor V (Culture), including some (e.g., M. I. 117: Highbrow tastes) with fairly strong associations. Peer-rating Factors I (Surgency or Extraversion) and III (Dependability or Conscientiousness) were more modestly associated with these 16PF and O-A Battery variables.

The variables listed in Table 4 were included in a series of stepwise multiple-regression analyses, and the results are summarized in Table 6. For each peer-rating criterion in turn, three regression analyses were carried out using, respectively, (1) those 16PF scales significantly associated with that criterion, (2) the significantly associated O-A Battery variables, and (3) the combined set of significant correlates from both the 16PF and the O-A Battery. Only the first four steps of each analysis are presented. These results, which have not been cross-validated, are presented purely for illustrative purposes; if the same equations were applied to new samples, the size of these multiple correlations would be considerably lower. In general, the results based upon the multiple-regression analyses follow the same pattern of differential predictability as those displayed in Table 4.

The findings presented so far, however, do not bear directly on the central problem under study, namely the comparative validity of Q data (16PF scales) versus T data (O-A Battery measures) in predicting the five peer-rating criteria. To answer this question, the factor scores based upon the 18 O-A Battery oblique factors were compared with the composite 16PF scores (Forms A and B combined) as multiple predictors of each peer-rating factor. The multiple correlations at each of the first 4 steps of these 10 stepwise regression analyses are presented in Table 7. These results suggest (1) that the O-A Battery measures were slightly superior to the 16PF scales in the prediction of peer-rating Factor V (Culture), (2) that the 16PF scales were slightly superior in predicting Factor III (Dependability or Conscientiousness), and (3) that the two sets of predictors were essentially equivalent in predicting the other three peer-rating factors. Across all five criteria there was no clear superiority of O-A Battery measures over 16PF scales.

Discussion

Although objective tests are more cumbersome and costly to administer than are questionnaires, they could well repay the increased investment in time and money if they provided substantially more valid measures of important individual differences. For example, question-

[2]The correlations are available from the first author.

[3]The zero-order correlations between each of the five peer-rating factors and the scale scores from the two separate forms of the 16PF, as well as from the composite scores (Forms A and B combined), are available from the first author, along with the analogous correlations based upon the 57 O-A Battery variables and those based upon factor scores from the 18 oblique O-A Battery factors.

Table 4
Significant Correlates of the Five Peer-Rating Factors
(N = 82)

	Peer-Rating Factor				
Instrument	I	II	III	IV	V
16PF					
F	.27		-.36		
G	-.29				-.34
H	.26				
I					.33
M	.25		-.31		.38
O				-.24	
O-A Battery					
MI 7			-.24		
MI 9					.23
MI 25a					.26
MI 87					-.27
MI 102					.26
MI 117					.47
MI 145a		.23			
MI 159a	-.33				-.30
MI 167a					.39
MI 206			-.25		
MI 219					.24
MI 237				.23	
MI 246	-.25		.23		-.31
MI 275			-.22		
MI 316		.25			
Factor 9			.22		
Factor 10	.38				.55
Factor 11					.29
Factor 15					.25
Factor 18			.24		

naires are often viewed as alarmingly vulnerable to individual differences in impression management (e.g., response sets, biases, or styles), while objective tests are ostensibly more immune to such potentially distorting effects. In fact, Cattell has explicitly defined objective tests as those on which " . . . the subject does not really (he [or she] may believe he [or she] does) know for certain in what way his [or her] behavior is being measured or what kinds of personality inference will be drawn from his [or her] test reactions" (Cattell & Warburton, 1967, p. 16); and he has argued that " . . . the real art of the psychologist . . . is to produce the kind of test which disguises (from the subject) what it measures!"(Cattell & Warburton, 1967, p. 35).

Unfortunately, as has been argued elsewhere (Goldberg, 1968), this definition shifts the criteria for establishing objectivity from the test materials and instructions to the introspections of

Table 5
Significant O-A Battery Correlates of the Five Peer-Rating Factors ($N = 82$)

Peer-Rating Factor and Correlation	*Master Index No. of Variable*	*Test No.*	*Test Name*	*Description of Variable*
I: Surgency				
−.33	159a	T19	Time Estimates for Everyday Tasks	greater variability of other-referent time estimates
−.25	246	T34	Authoritarianism: Personal Values	little submission
II: Agreeableness				
.25	316	T13	Criticalness of Self and Other Persons	more fluency concerning own, relative to other people's, personal characteristics
.23	145a	T17	Aspiration Level in Coding	higher responsiveness of aspiration level to performance
III: Dependability				
−.25	206	T37	Gottschaldt: Perceptual Analysis	greater accuracy
−.24	7	T2	Gestalt Closure	faster speed of closure
.23	246	T34	Authoritarianism: Personal Values	little submission
−.22	275	T30	Arithmetical Reasoning	higher total correct
IV: Neuroticism				
.23	237	T42	Mazes (Pencil)	faster speed on difficult, relative to easy, mazes
V: Culture				
.47	117	T27	Highbrow Tastes	more highbrow tastes
.39	167a	T9	Opinion Inventory X	better immediate memory
−.31	246	T34	Authoritarianism: Personal Values	little submission
−.30	159a	T19	Time Estimates for Everyday Tasks	greater variability of other-referent time estimates
−.27	87	T6	Reading Tempo and Memory for Interesting Material	better recall of emotionally interesting, relative to dry, material
.26	25a	T15	Memory for Emotional Words	better recall of emotional, relative to nonemotional, words
.26	102	T18	Aphorism Acceptance	greater tendency to consider remarks "obvious"
.24	219	T41	Honesty in Admitting Frailties	more frailties admitted
.23	9	T50	Ideomotor Tempo	faster tempo

Table 6

Results of the Stepwise Multiple Regression Analyses: Selected 16PF Scales (Forms A+B) and Selected O-A Battery Variables Which Were Significantly Correlated with the Peer-Rating Criteria (N = 82)

	Peer-Rating Factor									
	I		II		III		IV		V	
Test Step	Variable	R	Variable	R	Variable	R	Variable	R	Variable	R
16PF										
1	G	.29			F	.36	O	.24	M	.38
2	H	.37			M	.44			G	.45
3	M	.43							I	.47
4	F	.44								
O-A Battery										
1	159a	.33	316	.25	206	.25	237	.23	117	.47
2	246	.40	145a	.32	246	.31			167a	.57
3					7	.35			159a	.63
4					275	.40			25a	.65
Both										
1	159a	.33			F	.36	O	.24	117	.47
2	G	.44			M	.44	237	.33	167a	.57
3	M	.48			275	.48			159a	.63
4	H	.52			7	.51			M	.67

Note. Since these multiple correlations have not been cross-validated, their values should not be construed as applicable to new samples.

Table 7
Results of the Stepwise Multiple Regression Analyses: 16PF Scores (Forms A+B) vs. O-A Battery Factor Scores

	Peer-Rating Factor									
	I		II		III		IV		V	
Step	16PF	18 O-A	16PF	18 O-A	16PF	18 O-A	16PF	18 O-A	16PF	18 O-A
I. Total Sample (N = 82)										
1	.29	.38	.17	.20	.36	.24	.24	.14	.38	.55
2	.37	.41	.25	.28	.44	.34	.34	.19	.45	.61
3	.43	.43	.33	.32	.47	.41	.36	.23	.50	.67
4	.45	.45	.37	.34	.50	.44	.38	.25	.51	.69
II. Average Cross-Validity Coefficients (N = 41 in each of two subsamples)										
1	.19	.37	-.07	-.09	.29	.09	.14	-.05	.28	.56
2	.22	.26	-.12	-.07	.40	.10	.12	.03	.31	.54
3	.24	.25	-.02	-.12	.34	.13	.13	.03	.36	.53
4	.17	.18	-.13	-.06	.32	.16	.05	.01	.34	.57

each individual being tested. A test, by Cattell's definition, is objective if and only if the examinee does not know the tester's scoring rationale; consequently, for any given test and testing session, some of the examinees may be administered an objective test and others may not. Moreover, for any one test, some scoring keys will be objective (for some examinees) and other keys will not. Of even greater significance, Cattell's definition of objectivity places the concept outside the usual domain of scientific usefulness, since most psychologists are unwilling to concede knowledge of the thoughts of other individuals.

However, to the extent that such subjective reactions can be inferred, many of the 30 O-A Battery tests utilized in the present project would fail Cattell's definitional criterion. In fact, those tests that are the most objective by the usual standards (e.g., the maximum performance tasks) are precisely those for which the average person should be able to intuit the "way his [or her] behavior is being measured." For example, of these 30 O-A Battery tests, about half are maximum performance tests of various abilities (e.g., Gestalt closure, logical ability, arithmetical reasoning, maze performance, clerical accuracy, clerical speed). For these tests, the measure variable and the testing task (as well as the test name) are typically much the same.

Moreover, the remaining half of the test battery includes mostly short questionnaires, which presumably reflect the same response processes endemic to self-report data more generally. Indeed, Cattell has elevated some of these putative response sets to the level of objective test scores. For example, three such biases (response extremeness, acquiescence, and response inconsistency) are scored from a short personality questionnaire, and social desirability response set is scored from another similar questionnaire. Although it may be assumed that most people would not suspect the meaning of these particular variables, the average person might well know the meaning of many others. For example, a test called "Highbrow Tastes" to the psychologist and "Personal Tastes" to the examinee in-

cludes 20 questions, each followed by three alternatives, one of which is keyed as highbrow (e.g., Item 1: When playing cards with your friends, which game would you prefer? (a) Canasta, (b) Bridge, (c) Poker). The test score is the number of highbrow alternatives selected. A test called "Annoyability" to the psychologist and "Common Annoyances" to the examinee includes 25 annoying happenings (e.g., Item 1: People who keep their radios loud), to each of which the examinee responds by checking one of three alternatives (Very Annoying, Somewhat Annoying, or Not Annoying). The variable scored from this test is called "more susceptibility to annoyances." The reader must ask how much art was involved in producing the subterfuge behind those two objective tests.

Finally, some of the O-A Battery tests appear to be but shortened versions of rather traditional personality or attitude scales. As just one of many possible examples, a test called "Authoritarianism" to the psychologist and "Good and Bad Personal Values" (sic) to the examinee includes 16 items (e.g., Item 1: Respecting authority and being obedient to law is not as important in the modern world as in the time of our forefathers), each administered with an agree-disagree response format. The test score is called "little authoritarian submission." Although it is possible that most people may not know the name of the variable scored from this test, they should share this ignorance with their peers who are responding to such classic attitude scales as the California F-scale or Rokeach's Dogmatism scale—both of which ostensibly provide Q rather than T data.

Indeed, it can be assumed that far less effort was devoted to establishing the reliability of most O-A Battery test scores than is usual with modern personality scales; and consequently that their potential relationships with other variables could be severely attenuated. Some evidence for this conjecture comes from two sources: (1) the extraordinarily low values of the intercorrelations among the 57 O-A Battery measures; and (2) the urgent warnings by Cattell that relatively minor changes in test administration conditions tend to produce massive changes in O-A Battery test scores. For example, Cattell (1971) has pleaded for keeping the "test administration conditions as closely similar as possible to those of the several studies that have been done [in his laboratory] since, even with the utmost care, it is easy to lose that comparability and that invariance which we seek in science." This concern, which is reminiscent of similar comments made 20 years ago by many exponents of the "projective hypothesis," is seldom given voice in the debate over the relative advantages of objective tests as compared to questionnaire scales.

However, any such lack of robustness in O-A Battery test scores makes factor analysis of these variables a dubious venture. Since the intercorrelational pattern among these 57 O-A Battery variables is quite similar to that produced by correlating random normal deviates, Cattell's usual criterion for estimating the quality of factor rotation (the hyperplane count, defined as the percentage of factor loadings between $-.10$ and $+.10$) can be grossly misleading. For example, when the 57×57 O-A Battery correlation matrix from the present study is factored, the hyperplane count tends to be quite high, no matter how many factors are rotated and no matter what sort of rotational algorithm is employed.

Nonetheless, the proof of the objective test pudding clearly lies in the validating, and here another serious problem arises. In their objective test compendium, Cattell and Warburton (1967) have described over 400 tests, yielding more than 2,000 scores, some of which *might* be useful in any particular applied setting. In the absence of any explicit theoretical links between the test scores and criterion behaviors, however, an enormous number of people would be needed to discover and to cross-validate any potentially important predictive measures included in this set. Given the findings from the present study, whose results showed no clear superiority of O-A Battery measures over 16PF scores in predicting broad peer-rating criteria, it is not obvious that such an awesome test distillation process is yet warranted.

References

Allport, G. W., & Odbert, H. S. Trait-names: A psycho-lexical study. *Psychological Monographs.* 1936, *47* (1, Whole No. 211).

Buros, O. K. (Ed.), *Personality tests and reviews.* Highland Park, NJ: Gryphon Press, 1970.

Cattell, R. B. Confirmation and clarification of primary personality factors. *Psychometrika.* 1947, *12.* 197–220.

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

Cattell, R. B. Personal communication, September 13, 1971.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. *Handbook for the Sixteen Personality Factor Questionnaire (16PF).* Champaign, IL: IPAT, 1970.

Cattell, R. B., & Warburton, F. W. *Objective personality and motivation tests: A theoretical introduction and practical compendium.* Urbana, IL: University of Illinois Press, 1967.

Goldberg, L. R. Explorer on the run (Review of *Objective personality and motivation tests: A theoretical introduction and practical compendium* by R. B. Cattell & F. W. Warburton). *Contemporary Psychology.* 1968, *13.* 617–619.

Goldberg, L. R. A historical survey of personality scales and inventories. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 2). Palo Alto, CA: Science and Behavior Books, 1971.

Hundleby, J. D., Pawlik, K., & Cattell, R. B. *Personality factors in objective test devices: A critical integration of a quarter of a century's research.* San Diego, CA: Knapp, 1965.

Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology.* 1963, *66.* 574–583.

Norman, W. T., & Goldberg, L. R. Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology.* 1966, *4.* 681–691.

Tupes, E. C., & Christal, R. E. Recurrent personality factors based on trait ratings. Technical Report ASD-TR-61-97. Personnel Laboratory, Aeronautical Systems Division, Air Force Systems Command, United States Air Force. Lackland Air Force Base, Texas, May 1961.

Acknowledgments

This project was initially supported by Contract AF 41 (657)–269 from the Personnel Laboratory, Aeronautical Systems Division, Air Force Systems Command, Lackland Air Force Base, TX, and subsequently by Grant MH 12972 from the National Institute of Mental Health, U. S. Public Health Service.

Author's Address

Send requests for reprints or for additional tables to Lewis R. Goldberg, Institute for the Measurement of Personality, 1201 Oak Street, Eugene, OR 97401.