Equivalence or Chaos: A Hypothesis testing Factor Analysis Study of the Role of Behavior Rating, Questionnaire and Instrument Factors in Personality Structure Research[1]

K. Warner Schaie

University of Nebraska

A growing controversy seems to be in progress among researchers and theorists concerned with the definition of the basic elements of personality structure via methods of multivariate analysis. The controversy pertains to the presence or lack of equivalence of certain personality factors isolated from the questionnaire and behavior rating domains. It is the purpose of this paper to give a brief description of this controversy, to identify what are believed to be its central issues, to propose a strategy for their investigation and to describe an experiment which attempts to carry out this strategy.

Cattell (1957) proposes that there are at least fifteen identifiable personality factors which are identical or at least similar in psychological meaning in both the questionnaire and behavior rating domains. The factorial stability of these factors has been reasonably well established in a number of replications conducted in Cattell's laboratory. Although some recent analyses (Peterson, 1960; Norman, 1963) suggest that, at least for behavior ratings, a smaller number of orthogonal factors may suffice, it may be assumed that the

Journal of Consulting Psychology
1963, Vol. 27, No. 4, 350-357

SCALING THE SCALES:

USE OF EXPERT JUDGMENT IN IMPROVING THE VALIDITY OF QUESTIONNAIRE SCALES [1]

K. WARNER SCHAIE
University of Nebraska

It is hypothesized that the concurrent validity of questionnaires can be increased by the use of item weights obtained by expert scaling instead of using the conventional unit weights. Items from 11 scales of the High School Personality Quiz (HSPQ) were rewritten in a form permitting application of the constant sum method to the judgment of item weights. The scales were rated by 30 psychologists and ratio scales constructed for each factor scale. The HSPQ records of 43 delinquent girls were correlated with behavior ratings on 42 traits from Cattell's "normal trait sphere" as judged by the Ss' cottage parents. Validity coefficients are given for the HSPQ unit weight scores and for scores weighted by expert-derived ratio weights. Results show low magnitude increments in validity. Effects of scaling on the problem of agreement among questionnaire and rating data were also evaluated with similar results.

The traditional procedure in constructing questionnaire scales consists of defining an item pool and using empirical methods to derive subsets of items which are relevant to some observed or postulated dimension. Unit weights are usually assigned to all items because there is no information on relative item weights. Even when factor weights offering greater precision are available they are generally ignored for practical reasons or because it is argued that increments in precision due to differential item weights will be trivial.

The items selected for a given questionnaire scale on empirical grounds may in fact be the best items available but there is no assurance that the resulting unit weight scale has certain properties which are likely to maximize validity. An important property for this purpose would be the requirement that the questionnaire scale be a ratio scale. Factor analytic or other empirical procedures of scale construction can obviously provide no assurance of this property.

[1] A preliminary version of this paper was presented at the annual meeting of the Society for Multivariate Psychology, November 1962, at Chicago, Illinois. The author is grateful for the cooperation of his colleagues and graduate students who performed the expert scaling, and to the staff at the State Training School for Girls, Geneva, Nebraska, for their cooperation in the testing and behavior rating of the questionnaire respondents.

It may furthermore be argued that the ultimate criterion of the utility of a questionnaire will be found in its ability to predict traits or behaviors which are confirmable by direct observation. The psychologist in his role as an expert on human behavior may have relevant information on the relations between self-report and observable traits. It is therefore proposed that an attempt be made to utilize the intuitive processes and capitalize upon the training and sophistication of the psychologist by utilizing his judgment to improve the weights assigned to data derived from actuarial methods. As has been pointed out by Meehl (1954), this does *not* mean that the psychologist is a superior computer. It means that he has been programed with rules which are not yet explicit and which are therefore not available for direct implementation. Such rules would include that of rejecting data which are patently absurd although possible, or of assigning different weights or discounting weights for data which for computational efficiency have been handled as if they were of equal weight even though no such equality exists.

In spite of the quite popular feeling that there is a wide gulf between psychometric and psychophysical procedures, one may utilize the latter to advantage to maximize desirable scale properties. The constant sum method (Metfessel, 1947) provides a promising approach for obtaining ratio scales on more or less subjective psychological variables. The method has been applied mostly where physical standards were also available, but it has been found useful and reliable in experiments scaling such matters as preference for neckties (Dudek & Baker, 1956) and the association between colors and mood tones (Schaie, 1961).

It is now proposed to apply this constant sum method to the scaling of empirically derived personality scale items and to examine the effect of such scaling on the concurrent validity of the questionnaire scales in predicting trait ratings and factor scores based on such trait ratings.

PROCEDURE

The High School Personality Quiz (HSPQ) was developed by Cattell, Beloff, and Coan (1958) to provide a comprehensive personality description of adolescents by means of factored scales. The HSPQ contains 14 questionnaire scales each consisting of 10 items with unit weights. According to the authors, 11 of the 14 scales assess factors which have also been identified in the behavior rating domain. This study is therefore limited to the 11 scales for which validity data are available.

The items on the 11 HSPQ scales on Form A of the questionnaire (Factors, A, B, C, D, E, F, G, H, I, J, and O) were examined to determine whether they were in unidirectional form. The items which did not meet this criterion were then rewritten. For example, the item "At a picnic, would you rather spend some time: a. exploring the woods alone? b. playing around the campfire with the crowd?" was rewritten in the form "At a picnic I would rather spend some time playing around the campfire with the crowd than exploring the woods alone."

The rephrased items were arranged in a booklet with the 10 items of each scale on one page, appearing in the order in which the items appear in the questionnaire. These booklets were then handed to 30 psychologists (their qualifications being defined here as eligibility for election as APA associates) who were asked to scale items by means of the constant sum method.

The constant sum method requires the judge to indicate the relative magnitude of two stimuli by dividing 100 points between them. Although the procedure generally used in the constant sum method is patterned after the method of paired comparison there is no need for this particular model. In fact the large number of indirect estimates may introduce different judgmental biases depending on the dissimilarity of the stimuli which are to be compared. For this reason, as well as to keep the judgmental task within reasonable bounds, a modified technique using the method of the constant stimulus as a model was applied. This modification was originally applied to the scaling of line lengths (Baker & Dudek, 1957), where physical dimensions are available, but has also been found useful and reliable in the scaling of the association of colors and mood tones (Schaie, 1961).

The task of the expert judges was to compare Item 1 on each scale with each of the other nine items and to divide 100 points between each item pair assigning the larger number to the item rated more appropriate to the positive end of the factor dimension. As an example the instructions for Factor A (Cyclothymia versus Schizothymia) were as follows:

> Compare statement 1 with each of the following statements. Divide 100 points between each pair of statements to judge their relative position on the dimension "warm, sociable vs. stiff, aloof." Record only the number of points assigned to statement 1. The larger number of points should be assigned to the statement judged to be more "warm, sociable."

Scale values were obtained by totalling the points assigned to Item 1 in comparison with all the other items and the ratio of the sums of all possible points to the total for Item 1 was calculated for each of the compared items. To permit comparison with the unit weights of the original questionnaire scale values were next transformed to a 10-point scale. On practically all scales some item was rated as half as relevant or important as some other item with respect to the dimension on which it had been placed. For example, the item weights obtained for the scale for Factor A were 1.4, 1.0, 0.9, 0.9, 0.8, 0.7, 1.6, 1.1, 0.7, 0.9.

The new item weights were next applied to the rescoring of a set of HSPQ protocols obtained from a group of 43 girls who were residents of a training school for delinquent adolescents. These girls were also rated by their cottage parents on the 42 traits described by Cattell (1957) as providing an adequate sample of what he calls the "normal trait sphere." Furthermore, factor scores were computed from these trait ratings using Cattell's weights which yielded measures for 15 rating factors. Intercorrelations of the HSPQ measures, validity coefficients describing the relationship between the HSPQ and the trait ratings and rating factor scores, as well as multiple correlations between the questionnaire and criterion data, were then obtained to compare the effect of the new scores using the weights provided by the experts as contrasted to the original unit weights. The following section will discuss the results of these analyses in some detail.[2]

[2] Tables giving the ratio item weights for each of the 11 scales, means, standard deviations, scale intercorrelations, and multiple correlations with the criterion variables have been deposited with the American Documentation Institute. Order Document No. 7557 from ADI Auxiliary Publications Project,

TABLE 1

MEAN DIFFERENCES, CORRELATIONS, AND PROPORTION OF INDEPENDENT VARIANCE BETWEEN UNIT WEIGHTED AND RATIO WEIGHTED SCORES

	Factors										
	A	B	C	D	E	F	G	H	I	J	O
Mean difference	−.02	.08	.26	−.18	.18	.05	−.07	−.13	−.14	−.18	−.24
Correlation	.97	.95	.94	.96	.96	.97	.98	.99	.98	.95	.98
Proportion of independent variance	6%	10%	12%	8%	8%	6%	4%	2%	4%	10%	4%

RESULTS

The first question to be examined is the effect of the scaling upon the descriptive statistics of the HSPQ performance of our sample of questionnaire respondents. In evaluating differences between means the proper alternative to the null hypothesis is that the means for the scaled scores should be lower than the means for unit weight scores since unit weight scores should tend to overestimate scores of individuals rating low on a given scale. A difference significant at the .05 level of confidence appears only for Factor O.* However, keeping in mind the relatively small sample and the accompanying danger of Type II error, it should be reported that differences in the expected direction which were significant at the .10 level of confidence were found also for Factors D, H, I, and J. An inspection of score variability further indicates that variability increases as a result of scaling for all scales except Factors H and I.

Intercorrelations were also computed for the two score matrices and the Kolmogorov-Smirnow test of the difference between two correlation matrices (Schaie, 1958) was applied. It was found that the two matrices did not show more than chance deviation from one another and that the scaling therefore apparently had no effect upon the general structure of scale intercorrelations. This is desirable since the factor-analytic solution presumably has obtained the best set of items with respect to factorial structure and optimal estimation of factor scores.

Table 1 lists the mean differences between unit weight and ratio scaled scores, as well as the correlations and the proportions of independent variance available for each scale. The latter, ranging from 2% to 12%, represent the limit of the increment in precision which may be expected as a result of the scaling procedure. The magnitude of this increment is small, but it may be artificially restricted by requiring adherence to the original 10-point scale.

Our basic premise for this study was the hypothesis that the concurrent validity, i.e., the statements which can be made about the ability of the HSPQ to predict criterion behavior, would be increased by expert weighting of items. The next logical step therefore is an examination of changes in the validity coefficients. These are reported in Table 2. It may first of all be noted that the number of validity coefficients which may be interpreted as differing significantly from zero at the .05 level of confidence has increased after scaling. Sixty-eight coefficients significant by the above criterion were found for the unit-weight scores, while 77 such coefficients appeared after applying the scaled weights. Of the 81 coefficients reaching the .05 level of significance by either scoring method, 45 showed increase in magnitude after scaling, 18 retained the same magnitude, while 18 showed some decrease. No changes in magnitude, however, exceed 1 point in the first decimal place.

Questionnaire results are rarely interpreted scale by scale. More frequently use is made of the scale profile or regression equations

Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief Photoduplication Service, Library of Congress.

TABLE 2

BISERIAL CORRELATIONS BETWEEN HSPQ FORM A AND TRAIT RATINGS

Traits	HSPQ Factors										
	A	B	C	D	E	F	G	H	I	J	O
1. Considerate-inconsiderate	−11 −06	36* 41*	16 14	−24 −17	12 12	−05 −03	15 20	−07 −06	−20 −18	06 16	−08 −12
2. Calm-excitable	−20 −19	−12 −17	−23 −22	−16 −16	05 03	−19 −25	−17 −16	−02 −05	09 10	38* 41*	−04 −12
3. Energetic-tired	08 06	11 13	10 07	−06 −12	00 −04	34* 33*	−06 00	25 24	−25 −21	−21 −25	−23 −20
4. Quiet-noisy	−03 −02	13 12	−42* −36*	19 16	03 −11	−43 −49*	14 13	−39* −38*	13 17	28 34*	08 05
5. Patient-impatient	16 13	22 21	00 04	−16 −16	10 02	00 −05	19 22	05 08	−28 −25	06 10	−25 −27
6. Cheerful-solemn	01 02	−04 −05	04 −01	−12 −13	16 27	41* 40*	−24 −28	23 21	−19 −22	−25 −30	−07 00
7. Friendly-reserved	−03 −04	−28 −20	32* 35*	−24 −23	00 −09	55* 50*	03 06	17 16	−14 −17	−40* −42*	−25 −24
8. Meditative-unquestioning	22 22	29 24	03 04	−09 −16	−24 −31*	−30* −33*	49* 45*	−07 −01	17 16	11 22	−25 −31*
9. Cooperative-obstructive	−10 −03	23 28	40* 44	−03 −02	09 12	24 25	36* 41*	05 03	−27 −27	−09 −04	−18 −22
10. Happy-sad	−23 −24	·02 06	00 −04	−31* −36*	13 11	23 25	−04 −01	08 09	−06 −07	03 00	−10 −09
11. Sensitive-tough	06 15	33* 32*	12 21	−11 −04	13 05	−10 −15	40* 44*	−08 −09	−23 −21	26 32*	−12 −15*
12. Intelligent-stupid	−02 01	33* 26	−07 −06	42* 46*	−01 −02	04 04	31* 38*	−05 −04	−29 −26	−29 −28	09 04
13. Poised-flustered	−26 −35*	05 02	−25 −24	04 −01	−02 00	−21 −21	−12 −10	03 −01	16 21	10 10	03 −02
14. Tolerant-jealous	17 19	04 07	08 22	00 08	15 13	−07 −10	21 23	22 20	−28 −26	22 26	−28 −31*
15. Dominant-submissive	−10 −12	−04 00	00 −07	−07 −06	−29 −19	26 27	−25 −30*	00 00	17 13	−46* −44*	09 12
16. Relaxed-tense	−12 −09	−11 −09	−35* −25	−18 −12	08 00	01 −01	−05 02	09 05	21 21	17 22	08 05
17. Conventional-unconventional	−16 −12	22 19	19 26	17 18	04 00	−21 −20	25 26	05 05	17 19	14 17	−15 −16
18. Sociable-self-contained	10 13	−50* −46*	07 03	−19 −17	23 26	30* 27	−16 −17	−11 −15	−13 −17	−03 −02	29 31*
19. Trustful-suspicious	−10 −12	20 18	19 32	11 14	15 19	−07 −05	−13 −12	16 18	−17 −14	02 09	−06 −10
20. Self-effacing-egotistical	16 14	16 12	11 18	10 04	12 08	−04 −06	−03 −03	−17 −18	−04 01	19 30*	01 00
21. Conscientious-unscrupulous	18 21	−18 −19	16 21	−08 −15	24 22	04 03	26 24	−22 −22	−18 −18	09 14	−16 −20

TABLE 2—Continued

Traits	HSPQ Factors										
	A	B	C	D	E	F	G	H	I	J	O
22. Adventurous-	12	−12	48*	−10	−12	41*	14	54*	−38*	−34*	−30*
timid	15	−12	46*	−02	00	43*	15	54*	−41*	−45*	−29
23. Stable-	−02	34*	−08	−03	−03	−06	28	13	−29	−26	04
unstable	−07	34*	−02	−04	−05	−01	37*	13	−25	−19	−02
24. Persevering-	09	**04**	01	−34*	−21	01	28	03	−02	06	−17
quitting	13	**03**	04	−39*	−28	−01	30*	05	−05	12	−21
25. Modest-	−22	32*	−27	−02	**03**	−30*	08	−43	17	32	19
attention-seeking	−21	34*	−31	−09	**−04**	−26	09	−42*	20	40	16
26. Open-	23	−01	−15	03	26	−03	−16	−48*	−36*	33*	38*
defensive	26	03	−12	−01	26	−01	−10	−49*	−34*	37*	32*
27. Refined-	−18	15	−24	08	29	−20	−05	−42*	−11	06	23
crude	−17	06	−21	−01	23	−22	−03	−40*	−06	17	21
28. Imaginative-	−01	−39*	43*	00	−13	06	−09	30*	**−07**	−07	−37*
practical	−07	−44*	48*	05	−07	01	−17	33*	**−06**	−15	−32
29. Obedient-	−01	08	38*	−27	**03**	02	15	02	−30*	−07	−30*
disobedient	01	04	46*	−28	**07**	−02	12	05	−29	01	−34*
30. Adaptable-	−19	35*	11	−11	15	−21	18	16	−33*	−01	−12
inflexible	−17	34*	20	−09	21	−17	21	15	−28	04	−15
31. Responsible-	09	20	03	−15	−05	−10	**59***	−20	−16	−05	−12
irresponsible	16	16	06	−18	−06	−12	**61***	−16	−19	04	−15
32. Curious-	−07	−17	47*	00	−04	30*	06	17	−21	−41*	−22
incurious	−09	−12	37*	−03	09	36*	02	17	−20	−42*	−22
33. Talkative-	09	03	15	06	06	37*	−07	16	−36*	−36*	−02
silent	08	01	11	09	24	39*	−09	16	−37*	−48*	03
34. Carefree-	−06	−15	30*	05	−06	44*	−34*	25	−16	08	−18
anxious	−09	−13	32*	05	06	48*	−33*	25	−20	−01	−16
35. Tasteful-	18	−24	20	−02	−09	27	18	−08	**−05**	−20	−17
inartistic	22	−28	21	−03	−07	20	16	−06	**−07**	−12	−18
36. Resourceful-	−15	−11	−12	39*	23	**−02**	−05	−02	−42*	−09	23
baffled	−17	−11	−09	44*	23	**05**	03	−01	−42*	−13	18
37. Independent-	15	09	08	−02	08	07	08	09	−37*	00	−04
dependent	17	07	12	01	09	08	12	12	−38*	00	−08
38. Adult-	03	17	03	33*	−10	07	28	13	−29	−39*	08
naive	00	14	04	27	−07	13	35*	13	−22	−35*	03
39. Orderly-	−06	02	−23	−29	−01	−33*	**15**	−30*	−02	15	18
disorderly	05	−05	−21	−30*	−08	−35*	**12**	−29	−04	25	14
40. Easygoing-	−15	−15	39*	−25	24	00	−04	**02**	−45*	15	−12
irritable	−08	−09	35*	−16	28	01	00	**01**	−51*	16	−18
41. Expressive-	03	−23	04	−16	**10**	16	05	**−06**	01	−14	−03
secretive	01	−20	08	−15	**11**	12	04	**−05**	02	−13	−03
42. Brave-	23	−04	**27**	05	10	20	19	−06	−08	11	−25
complaining	29	−02	**32***	01	08	18	18	−06	−08	16	−28

Note.—The top figure in each row is for the unit weighted scores, the bottom figure is for the ratio weighted scores. Figures in bold face indicate correlations with the factor on which a given trait is hypothesized to load (Cattell, 1957).
* $p < .05$.

based on known relations between the scales and criterion measures. It was therefore felt to be appropriate to compute multiple correlations between the HSPQ and each of the 42 trait rating criterion variables. To avoid maximizing chance and artifactual relationships in the sample use was made of a technique described by Horst and Smith (1950) as programed for the Burroughs 205 computer by Schaie (1963). This is an iterative method which selects the largest validity coefficient as the first estimate of R, subtracts out the product of this value with the appropriate vector of the predictor matrix, and then selects the next largest remaining validity coefficient to be added. After each cycle the multiple R is corrected for shrinkage by means of Wherry's formula. Whenever the corrected R no longer increases or when it decreases, then the multiple R obtained at the end of the previous cycle is accepted as the best estimate of R and further iteration is ceased.

Of the 42 regression equations computed,

TABLE 3

CORRELATIONS BETWEEN HSPQ FORM A AND FACTOR SCORES DERIVED FROM TRAIT RATINGS

Trait Rating Factors	HSPQ Factors										
	A	B	C	D	E	F	G	H	I	J	O
A Cyclothymia versus	**−07**	−02	15	−17	17	06	−08	03	−24	05	−03
schizothymia	**−06**	−01	21	−14	22	05	−06	02	−23	09	−06
B Intelligence	06	**22**	−12	03	−22	−03	38*	−11	02	−13	−07
	09	**20**	−10	00	−23	−03	40*	−09	02	−08	−10
C Ego strength versus	07	00	**−04**	−10	06	00	09	00	−11	15	−12
proneness to neuroticism	07	−02	**02**	−13	04	−01	12	00	−10	18	−17
D Excitability versus	−07	−06	−04	**14**	−05	04	−11	04	08	−25	15
insecurity	−09	−06	−09	**16**	−01	05	−11	03	07	−31*	18
E Dominance versus	−01	−28	05	03	**−08**	23	−15	06	01	−23	12
submissiveness	−02	−26	01	04	**−02**	25	−16	05	−02	−29	13
F Surgency versus	−12	11	16	−08	17	**16**	−08	28	−27	−07	−13
desurgency	−12	08	19	−06	24	**19**	−05	27	−25	−10	−14
G Super ego strength	02	11	−04	−21	03	−13	**26**	−14	−08	14	−08
	08	07	−01	−22	00	−15	**27**	−12	−09	22	−12
H Parmia versus threctia	−06	−23	39*	−14	06	41*	−06	**22**	−31*	−26	−07
	−06	−17	36*	−09	19	43*	−04	**21**	−35*	−33*	−07
I Premsia versus harria	−19	−08	13	03	−12	−14	05	08	**25**	−02	−03
	−20	−12	12	07	−10	−16	00	06	**26**	−01	−01
J Coasthenia	01	−04	−29	−14	−11	−44*	06	−17	33*	**46***	10
	02	−06	−29	−12	−17	−47*	02	−15	32*	**49***	08
K Contention versus	10	−11	08	−05	09	12	09	−11	−12	−04	−15
abcultion	11	−14	10	−06	10	10	08	−10	−11	−02	−17
L Protension versus	02	−04	−30*	10	−09	−19	−12	−19	18	03	24
inner relaxation	−01	−08	−39*	06	−13	−18	−14	−18	18	−01	26
M Autia versus	−04	−20	−14	−01	−15	05	−04	07	00	−05	03
Praxernia	−08	−24	−12	−01	−12	03	−03	07	−01	−06	00
N Shrewdness versus	14	17	−08	24	20	00	09	−23	−25	01	14
naivete	17	10	−09	22	19	01	13	−23	−21	08	10
O Guilt proneness versus	−01	−03	05	−12	−25	−18	11	−15	03	03	**14**
confidence	00	−01	−01	−10	−24	−19	08	−14	−01	02	**15**

Note.—The top figure in each row is for the unit weighted HSPQ scores, the bottom figure is for the scores ratio weighted by the Constant Sum Method. Figures in bold face indicate correlations between matching factors in the behavior rating and questionnaire domains.

* $p < .05$.

28 resulted in increased and 14 in decreased corrected multiple *R*s. On the average, scaling raised multiple *R*s by .037 and corrected multiple *R*s by .052. Twenty of the multiple correlations based on ratio weighted scores showed an increase in prediction significant at or beyond the .05 level of confidence. Carrying out the iterative process until all predictors had been assigned a weight changed individual coefficients but did not affect overall results. The reported shifts are not large, but they do seem to indicate a noteworthy trend.

EFFECTS OF SCALING ON CROSS-MEDIUM FACTOR MATCHING

An interesting extension of our experiment leads to the question whether it can contribute some clarification to the controversy regarding the matching of L (behavior rating) and Q (questionnaire) data (Becker, 1960; Cattell, 1961). Specifically we are concerned with the question of what if any effect the scaling will have in reducing the effects of such variables as the instrument and perturbance factors specified by Cattell to account for the lack of congruence between L and Q data. A more thorough approach to this problem would require factor analytic studies, which for the unscaled data have been reported elsewhere (Schaie, 1962), but which were not attempted in the present context. We shall be concerned here primarily with an examination of the effects of scaling on the zero-order correlations. If instrument and perturbance factors in fact conceal similarities in factor structure, then the increment in validity obtained by expert scaling may be a function of the decrease of effects of these semiartifactual components.

To study these matters, validity coefficients were computed between both scaled and unscaled HSPQ scores and Rating Factor scores which were obtained by combining trait ratings according to the factor structure reported by Cattell (1957). These coefficients are reported in Table 3. Matching between factor scores from the R and Q media is obtained at the .05 level of significance only for Factor J (Coasthenia). Although none of the coefficients for Factor G (Ego strength) reach significance, the matching loadings are highest. In terms of our significance criterion, mismatchings occur between C_q and Factors H_l and L_l; between Factor F_q and Factors H_l and J_l; between Factor G_q and Factor B_l; between Factor I_q and Factors H_l and J_l; and between Factor J_q and Factors D_l and H_l. What is the effect of the scaling on these relationships? In our sample it is restricted to raising most of the significant coefficients very slightly and in introducing two additional mismatches.

CONCLUSIONS

It is apparent from these results that there seems to be a consistent trend for an increase in the concurrent validity of the HSPQ scales as well as their combination after rescoring with expert-derived ratio weights. It is also apparent that the magnitude of the increases for our sample is not very great and the legitimate question arises whether the increment is worth the extra labor involved. It should be kept in mind, however, that our sample of questionnaire respondent was quite limited and that the criterion variables were certainly fallible. Also, we do not know what effect the rewriting of the questionnaire items had in modifying the judges' decisions. The results with respect to the cross-medium factor matching question are further limited by the fact that we used Cattell's factor weights for adults in computing factor rating scores. The factor structure of rating data on adolescents may, however, be somewhat different than that for adults, resulting in possible erroneous conclusions from our data. It would seem desirable, therefore, to conduct further studies similar to the one described here under more favorable conditions, since it appears that expert judgment can indeed increase the validity of questionnaires. The task then remains to investigate the experimental conditions under which the gain from expert judgment will be sufficiently large to merit practical application.

REFERENCES

BAKER, KATHERINE E., & DUDEK, F. J. Scaling line-lengths with modification of the constant-sum method. *Amer. J. Psychol.*, 1957, 70, 81-86.

BECKER, W. C. The matching of behavior rating and questionnaire factors. *Psychol. Bull.*, 1960, **57**, 201–212.
CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.
CATTELL, R. B. Theory of situational, instrument, second order, and refraction factors in personality structure research. *Psychol. Bull.*, 1961, **58**, 160–174.
CATTELL, R. B., BELOFF, H., & COAN, R. W. *Handbook for the High School Personality Questionnaire*. Champaign, Ill.: Institute for Ability and Personality Testing, 1958.
DUDEK, F. J., & BAKER, KATHERINE E. The constant-sum method applied to scaling subjective dimensions. *Amer. J. Psychol.*, 1956, 69, 616–624.
HORST, P., & SMITH, S. The discrimination of two racial samples. *Psychometrika*, 1950, **15**, 271–289.
MEEHL, P. E. *Clinical versus statistical prediction*. Minneapolis: Univer. of Minnesota Press, 1954.
METFESSEL, M. A proposal for quantitative reporting of comparative judgments. *J. Psychol.*, 1947, **24**, 229–235.
SCHAIE, K. W. Tests of hypotheses about differences between two intercorrelation matrices. *J. exp. Educ.*, 1958, **26**, 241–245.
SCHAIE, K. W. Scaling the association between colors and mood tones. *Amer. J. Psychol.*, 1961, **74**, 266–273.
SCHAIE, K. W. On the equivalence of questionnaire and rating data. *Psychol. Rep.*, 1962, **10**, 521–522.
SCHAIE, K. W. Multivariate prediction programs for the Burroughs 205 computer. *Behav. Sci.*, 1963, **8**, 170–172.

(Received December 28, 1962)