Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 294-302

CONCEPTUAL ANALYSIS OF PSYCHOLOGICAL TEST SCORES AND OTHER DIAGNOSTIC VARIABLES [1]

HARRISON G. GOUGH

University of California, Berkeley

For the clinician, the validity of a test or assessment technique resides in the range and structural clarity of the information it provides him about the individual client with whom he works. 2 devices, for example, might be equal in forecasting a particular criterion, yet differ widely in their personological implicatons. Analysis of this problem permits specification of 3 levels of evaluation: primary, secondary, and tertiary. The conceptual model defined by these levels would appear to incorporate the kind of information which the diagnostician desires, and which indeed he must have if he is to function in an insightful and fully professional manner.

The purpose of this paper is to offer a point of view concerning the meaning of measurement in psychology. Attention is centered on the use or application of such measures, and what they tell the interpreter about the individual who has been tested. Significant prior discussions of the validity issue in testing and diagnosis have sought to classify tests according to the criteria employed in their construction and evaluation (cf. Cronbach & Meehl, 1955), and to specify ways in which discriminations may be sharpened (cf. Campbell & Fiske, 1959). These emphases are important, but nevertheless do not touch on all of the significant facets of meaning subsumed under the concept of validity. The intention here is to present a different perspective, one which stresses the implications of any scale or variable when it is brought to bear upon the analysis of the individual case.

From this perspective, the practitioner in testing seeks variables which permit individuated descriptions of the subject who has been tested, forecasts of what he will say or do, and characterizations of the way in which others will react to him. The greater the range of such information, and the more accurate its specification, the greater the value of the instrument which produced it.

An adequate theoretical position, following this, is one which points clearly to the kind of information which the practitioner must have if he is to employ his diagnostic tools in a responsible and professional manner. It should also provide help in structuring and clarifying this information, so that he may readily see what it is that is known, and what still needs to be known.

Two assumptions in this formulation are (*a*) that the purpose of the test is to assess and/or forecast significant nontest behavior, and (*b*) that the test is intended ultimately for interpretation and analysis of the individual case. A third principle is that the organization and application of this knowledge can only be realized through the endeavors of the trained, professional practitioner.

Our problem in defining criteria of validity and meaning is therefore to keep in mind the needs of the clinician who will interpret the tests, and the individual case which provides the fundamental context for their application. This task is not easy, and it is not expected that all readers will agree with the solution attempted in this paper. It may even be that each practitioner must solve these issues for himself, or try to solve them, and that no single proposal can win general acceptance. The only warrant for the present effort, if this is true, is that even in disagreement the reader may be stimulated to new considerations and perhaps to a more logical and clinically meaningful personal perspective.

The Conceptual Model

Let us begin with a brief formulation of the conceptual model to be proposed, and after

[1] A preliminary version of this paper was presented in a symposium on "Approaches to Integrated Concepts of Personality Theory and Assessment," New York Society for Projective Techniques, New York, June 6, 1964.

that move to more extended discussion of its components and to illustrative examples. This model is organized around three stages in the evaluative processes, that is, three foci of understanding which the interpreter must fully comprehend if he is to achieve an adequate conceptualization of a diagnostic variable. The first of these emphases may be designated the *primary evaluation*. The task here is to determine what criteria are principally relevant to the test, how well it predicts what it seeks to predict, measures what it purports to measure, or defines what it is intended to define. Most of what is said about test validity in textbooks and manuals may be classified under this first heading.

Secondary evaluation seeks to discover the psychological basis of measurement, to specify and clarify the meaning of that which is is measured. A scale may forecast some important nontest behavior, such as the likelihood of improvement in psychotherapy, and thus meet the requirements of our primary evaluation; but the clinician needs to know more than this, he must know what it is that the scale reflects that leads to this favorable potentiality. The task is to uncover and hence illuminate the underlying psychological dimensionality that is inherent in any test or measure possessing primary utility. When we return for specific consideration of this topic I shall attempt to enumerate explicit steps which can be taken in pursuit of this psychological (to be distinguished from psychometric) understanding.

Tertiary evaluation is perhaps more difficult to define than the other two concepts. It is concerned with the justification for developing a particular measure, or for calling attention to a measure. Part of this justification will come from the intrinsic significance of the primary aim of measurement, and part may come from the range of implications delineated in the secondary analysis. But additional and possibly even greater significance may come from the spectrum of life settings, beyond any envisaged under its primary validity, for which the technique has predictive and explanatory relevance.

An ability test might be developed with a primary goal of predicting success in school; on analysis, it is found that it does this satisfactorily, but perhaps no more so than many other such tests. Under secondary evaluation it is discovered that the key psychological variable which seems to be involved is the ability to reorganize and recombine perceptions and experiences, rather than the memory for facts and events which predominates in other devices. Our interest in this new test might now be heightening. But suppose that on the tertiary level we find, contrary to nearly all research with "ordinary" scholastic aptitude tests, that our new device appears to relate significantly to indices of creative and original endeavor, and that its forecasting efficiency becomes better and better as we move farther and farther away from the explicitly academic criteria in whose behalf it was initially constructed.

This kind of evidence is just what one seeks in the tertiary analysis; it is the kind of evidence which arouses the special interest of the clinician and which justifies his paying special attention to the instrument which possesses it.

PRIMARY EVALUATION

It might be useful to link the discussions of each of the three stages of analysis by offering observations of a single diagnostic variable under each heading. For this purpose I should like to utilize the So or "socialization" scale of the *California Psychological Inventory* (Gough, 1957). The CPI is a true-false objective inventory scaled for "folk concepts," that is, variables used for the description and analysis of personality in everyday life and in social interaction. It is theorized that such folk concepts, viewed as emergents from interpersonal behavior, have a kind of immediate meaningfulness and universal relevance which enhance their attractiveness as diagnostic concepts. Hopefully, diagnoses and forecasts of social behavior, if mediated by such concepts, will be more accurate and dependable than forecasts arrived at by way of other formulations.

The system of concepts defined by the 18 scales in the inventory should also be mentioned. Its main property is that it is an "open" system—it can change and grow by the addition or substitution of scales, and if need be can be reduced by the elimination

of measures. Its purpose is to reflect social behavior, and in so doing to include a sufficient number of variables so that all major forms of such behavior can be forecast either by one scale or a combination or pattern of scales.

Within the full set of 18 variables there are certain natural clusterings, one of which refers to the domain of interpersonal values and intrapersonal controls. Three scales in particular bear on these issues: responsibility, socialization, and self-control. Although substantially intercorrelated, each is addressed to a different facet of the value complex: responsibility emphasizes the degree to which values and controls are conceptualized and understood; socialization emphasizes the degree to which they are internalized and made operational in the life of the individual; and self-control stresses the degree to which the individual approves of and espouses such regulatory dispositions.

Given these definitions, the primary validational task of the socialization scale is to locate individuals and groups along a continuum of asocial to social behavior, and to forecast the likelihood that any person will transgress whatever dividing line his own culture interposes between these two poles of the continuum. The phrase "his own culture" is used intentionally, because the folk concept theoretical basis of the inventory requires that its validity be demonstrated in other cultures than the one in which the sale is developed.

The So scale was first introduced in 1952 (Gough & Peterson, 1952). In 1960 a report was published (Gough, 1960b) which surveyed validational evidence obtained in the preceding 8-year period. For males, 25 samples involving 10,296 cases were considered, ranging from nominated "best citizens," through occupational samples of varying kinds, disciplinary problems, and county jail inmates, to incarcerated delinquents and felons. The biserial correlation for the dichotomy of more- versus less-socialized was +.73. For females, 16 samples totaling 10,560 subjects were studied, covering the same continuum of socialization; the biserial correlation here was +.78.

Primary validation evidence for classification of individuals along the full extent of the socialization continuum would appear adequate. The next question pertains to differentiation within zones or regions of the continuum. An interesting finding with respect to this question comes from the work of Vincent (1961) on unmarried mothers; for a sample of 232 women the biserial correlation between the So scale and the dichotomy of one illegitimate pregnancy versus two or more was +.83. Highly significant differences between recidivists and first offenders were found in studies of reformatory inmates (Donald, 1955) and juvenile offenders (Peterson, Quay, & Anderson, 1959).

Cross-cultural validation has been equally encouraging. For a sample of 203 institutionalized delinquents in India, tested in Hindi and Punjabi (Gough & Sandhu, 1964), the So scale correlated +.73 with court-assigned classifications of the severity of the offense. Just now we are reviewing data from Austria, Costa Rica, France, Germany, Italy, Japan, Puerto Rico, South Africa, and Switzerland, comparing delinquent versus nondelinquent samples of males and females. Nine countries are involved in this survey, and translations of the So scale into six languages: Afrikaans, French, German, Italian, Japanese, and Spanish. In every comparison the So scale has differentiated significantly between delinquents and nondelinquents; and, considering all samples, in no instances does the average for any sample of delinquents (no matter from which country or place) equal or exceed the lowest average score observed among the samples of nondelinquents.

There is also evidence from a study in progress by Ernest Wenk and the author on the prediction of parole outcomes. For a sample of 295 parolees followed for 3 years on parole it was found that 165 men were successful and 130 were violators. The So scale means for these two samples were 26.84 and 24.77; this difference, although not large, yields a t ratio of 3.33, which is significant beyond the .01 level of confidence.

The evidence just reviewed for the So scale is the kind of evidence we customarily expect on validity, and it is the kind of evidence which we must establish in our primary evaluation. The evaluation is primary because it comes first, and because there is no reason

to follow a measure on into the realms of secondary and tertiary evaluation unless we are reasonably satisfied that its primary utility is established.

SECONDARY EVALUATION

We are now ready to undertake a secondary evaluation of the So scale. Our aim here, to repeat, is to clarify the basis of measurement, that is, to determine what it is that the scale reveals about a person beyond the fact that he will probably behave in a more- or less-socialized manner. The four steps enumerated below are offered as aids in the accomplishing of this conceptual analysis.

Step 1. Review of the development of the measure, the procedures and samples used in its construction, its theoretical presuppositions and bases.

Step 2. Analysis of the components of the measure, its items, stimulus materials, and content.

Step 3. Determination of the relationships between the measure and (*a*) other measures already known and conceptualized, and (*b*) variables of self-evident importance such as sex, age, status, etc.

Step 4. Specification of the characterological and personological dispositions of individuals who obtain scores defined by the measure itself as diagnostically significant.

A primary evaluation will already have provided some of the information envisaged under Step 1, but other matters await consideration. Perhaps the most important of these is the conceptual basis of the measure. The So scale was initially conceived in the context of role theory. The less-socialized person was hypothesized to be less skillful in sensing and interpreting subtle and covert cues in social interaction, and hence less likely to evolve dependable and veridical internalized systems of control. Indirect evidence of such role-taking disability would be found in rule-breaking and rule-violating behavior, and direct evidence would be found in tests of the accuracy of role perspectives and interpersonal diagnosis.

In the construction of So, samples ranging from seriously delinquent through moderately delinquent to conventionally socialized were used, and item studies undertaken so as to identify those possessing validity for this continuum. The assumption here, the reader will recognize, is that one's manifest location along the socialization continuum will on the average provide a valid estimate of one's role-taking proficiency. Whether or not this assumption is true or partly true, a more direct test of the hypothesis is clearly needed.

To date, the most thorough check of this sort is found in a paper in 1957 by Reed and Cuadra. They studied 204 student nurses in a neuropsychiatric hospital. Each nurse described herself on the Gough Adjective Check List (ACL) (Gough, 1960a; Gough & Heilbrun, 1965) and then the other three nurses in a four-member group to which each was assigned. Next, each nurse attempted to guess how she would be described by her group. A point was earned for a predicted adjective if two of the three peers had, in fact, checked this word, and a total score was defined as the sum of these points. This score is therefore representative of the accuracy with which an individual senses the reactions of others to her, that is, of the degree to which she can take the role of the other and look upon self as an object. The correlation between the So scale and this social sensitivity index was +.41.

A related task would be to ask the individual to guess how another person would describe himself. In an unpublished study at San Quentin Prison in California, Bela Baker had inmates guess how their cellmates had answered the ACL; the accuracy of these estimates was significantly correlated with the inmates' own scores on the So scale.

Recapitulating briefly the information under Step 1, we can say that the scale may be conceptualized from the role-taking perspective and that there is both direct and indirect evidence for its validity from this perspective.

Step 2 takes us into the content of the scale. Here we find that its 54 items are about equally distributed over two kinds of content. The first of these more or less directly embodies role-taking ideas, and may be instanced by these items: "Before I do something I try to consider how my friends will react to it," "I often think about how I look and what impression I am making upon others," and "I find it easy to 'drop' or 'break with' a friend."

The other group of items has more immediate relevance to deviant or rule-breaking behavior, this being justified by the aforemen-

tioned hypothesis that asocial deviation is itself an indicator of the role-taking disability. Examples of this second type of item are: "I have often gone against my parents' wishes," and "If the pay was right I would like to travel with a circus or carnival."

The task for the practitioner in this second step is to internalize the content of the scale, and by virtue of this intimate familiarity to enhance the insightfulness of the psychodynamic formulation which is being evolved. The clinician who has studied and restudied the Rorschach inkblots and the cards of the Thematic Apperception Test need not be told that this second step, the personal, intense, apperceptive, and empathic perusal of the stimulus materials included in the scale or instrument is one of compelling importance and significance. It is unfortunate indeed that this step, invariably taken by projective testers, is only occasionally taken by psychologists in the ranks of objective or structured testing.

Step 3 concerns the relationships of the measure to other measures and to major categories of interpersonal variation. Perhaps the first question that should be asked of any test or scale being studied is "does it show a sex difference?" and the second "is it correlated with intelligence?" Among nondelinquent samples, both in the United States and elsewhere, the So scale does reveal a consistent sex difference of about half a standard deviation, with women scoring higher. One is tempted to say "as would be expected," as many would assert that women are in fact more law-abiding, more sensitive and perceptive, and more highly refined and elaborated in their role-taking. However, when attention is directed to delinquent and institutionalized samples, women tend to score a bit lower than their male counterparts.

Role-taking skill, if that is the underlying variable in the So scale, is probably correlated with intellect, but one would nonetheless hope that a personological measure could be defined which would be free of this particular component. The evidence seems to suggest that So is not correlated with intellectual ability. In seven or eight samples of fairly good size, varying with respect to age, educational level, etc., the correlations fall within a band of $-.14$ to $+.10$, with a median of close to .00.

Although So is unrelated to intelligence, it is related to operant conditioning under social reinforcement. Sarbin, Allen, and Rutherford (1965) compared 30 chronically delinquent boys with 30 nondelinquents, matched on age, intelligence, and occupation of parents, seeking evidence on the social responsiveness of delinquents to verbal conditioning. On two conditioning trials the delinquents did not differ significantly from the controls, but when higher-scoring delinquents on the So scale were compared with lower-scoring delinquents, significant differences in conditioning (favoring the higher scorers) were obtained; the same findings were observed among the controls, where those with higher scores on So conditioned more readily in the social learning task than those with lower scores.

What about other variables? Age norms would be of great interest, but unfortunately are not available. There is a hint, however, that the So scale rises more rapidly in adolescence and approaches the adult level earlier than do scores on the CPI scales for social presence, dominance, and self-acceptance.

Social status is another variable of interest. In several high schools the So scale yielded a median coefficient of $+.11$ with the Home Index (Gough, 1949), a measure of the socioeconomic status of the home and family background. Race differences have also been investigated, by Donald (1955) and Peterson et al. (1959) in the studies previously mentioned; in both instances Negro and white delinquents were *not* differentiated on the So scale.

Social desirability as a basis for response to the So scale might also be a source of worry. In a sample of 295 males, the correlation between So and Edwards' 39-item Social Desirability scale (Edwards, 1957) was $+.21$. In a sample of 152 males So correlated $+.01$ with Barron's scale for ego strength (Barron, 1953), $+.11$ with Welsh's anxiety index (Welsh, 1952), and $+.22$ and $-.23$ with the *K* and Psychopathic Deviate scales of the MMPI (Hathaway & McKinley, 1943).

A great deal of information of this sort is offered in the CPI *Manual* (Gough, 1957), and may be consulted there by anyone wish-

ing to push further into this third category. What has been presented is believed to be representative, and would seem to justify the conclusion that the So scale is a relatively independent variable, not contaminated by or unduly influenced by such factors as social desirability, status, anxiety, race, intellectual ability, or ego strength. Hence, it is not to be "explained" or "explained away" by such factors, but must be dealt with in its own right.

This leads us to Step 4, and a consideration of the immediate and direct personological implications of the scale. Although the previous steps are of interest to the clinician, and in fact vital if he wishes to reach a sophisticated and professional level of insight concerning any measure, it is the fourth step which comes closest to a direct delineation of the psychological meaning of a variable.

There are different ways of developing the information called for by this fourth step. One is the time-tested method of the individual clinician, who notes over the years of his practice what it is that characterizes patients scoring high or low on a particular test index—their defenses, their strengths, attitudes, self-conceptions, and stylistic predilections. The problem in this approach is that, even if valid, it takes years of patient accumulation of evidence before any sort of pattern begins to emerge. The need is for a faster and more efficient technique which retains the essential validity of the clinician's observations but shortens the time necessary to obtain them and to process them.

The personality assessment method, as practiced at the Institute of Personality Assessment and Research in Berkeley and at other centers, is exceedingly well adapted to do just this. Subjects come to the Institute in small numbers, and are studied intensively for 3 or 4 days by a panel of from 5 to 10 psychologists. Trait ratings, interviewers' formulations, Q-sort appraisals, and adjective check list descriptions are systematically gathered on all assessees. These sources of observational and diagnostic data may then be related back to the test scores of the assessees. The process is perhaps most clearly illustrated by ACL analyses, and only ACL data will therefore be drawn on in the discussion to follow.

Suppose 10 observers in assessment each complete a 300-word Gough Adjective Check List on an assessee. Any trait or quality which is above threshold in the eyes of the observer is checked, and any which is not is left blank. A convenient composite can be derived from these 10 individual descriptions simply by counting the number of times a word is checked, and treating that total as a score. Thus, if an assessee Adams is checked by all 10 observers as "alert" his score on this interpersonal quality or trait is 10; if 2 observers check him as "blustery" his score is 2; and if no one checks him as "charming" his score is zero. Adams will in this way be assigned 300 scores, 1 for each adjective.

If a sample of assessees has been described by the same panel of observers, these adjectival scores may be treated correlationally. A test variable, for example, the So scale, may be correlated with each of the 300 words in this sample of assessees. Adjectives showing significant and positive correlations are those which tend to be used to characterize high scorers on the scale, and hence afford a conceptual starting point for a personological sketch of the high scorer. Likewise with adjectives showing significant and negative correlations with the scale: those are the words that are in fact differentially used to characterize low scorers and they afford a valid starting point for a personological formulation of the low scorer. I believe that this procedure is intrinsically valid, and that it offers an efficient and powerful method for the psychological analysis of any variable that can be dichotomized or quantified. One thinks first of scales or single scores, such as for So, but the method is equally applicable to patterns, configurations, regression equations, or any other method of classifying, ranking, or calibrating individuals.

This method of analysis has been applied five or six times to the So scale. I should like to draw on four of these analyses, one based on a sample of 295 adult males, a second on a sample of 80 university graduate students, a third on a sample of 51 college seniors, and the fourth a sample of 100 military personnel studied by Reed and Cuadra. The first two samples are male, the third female, and the fourth includes both males and females; in

three samples the observations were contributed by psychologists who had studied the assessees, and in the last the descriptions were furnished by peers. The purpose in using all four samples is to overcome any limitations which might attach to a particular setting, use of professional versus peer evaluations, and to sex differences.

From these four independent analyses, those adjectives correlating in the same direction with the So scale in all four instances, and at a statistically significant level in at least two are presented below:

A. Adjectives used significantly more often to describe high scorers on the So scale.

calm	moderate	steady
considerate	modest	trusting
conventional	obliging	unassuming
cooperative	patient	
helpful	peaceable	

B. Adjectives used significantly more often to describe low scorers on the So scale.

affected	disorderly	irritable
arrogant	dissatisfied	rebellious
conceited	headstrong	restless
cynical	impatient	self-centered
defensive	impulsive	wary

A comment should be interjected here. Some psychologists, perhaps not given to thinking clinically, seem to be more bothered than helped by such a list of attributes. Such persons, one fears, are beyond the reach of the theoretical position being advocated in this paper. These adjectival clusters represent only the elements of clinical description, that is to say, protocol observations. From this starting point it is the interpreter's task to evolve an insightful diagnostic portrait of the high and low scorer on the scale, to render this evidence into an integrated formulation relevant for practice and understanding. The achievement of such a formulation requires all of the creativity that the interpreter can muster. But whatever its difficulty, the task must be attempted, for the elaboration of a psychological dynamic of the test variable is at the very heart of the conceptual analysis which is being illustrated.

TERTIARY EVALUATION

We come now to the third perspective, that of tertiary analysis. Is there any reason for paying attention to a scale or variable if one is not interested in its primary validational focus, that is, with respect to the So scale, in the problem of asocial versus socialized behavior?

Let us say that we grant that So can forecast asocial behavior with a surprisingly high degree of accuracy, and that it rests on a rather interesting theoretical and personological basis. Let us grant further that it is free of nuisance correlations with such variables as response sets, intellectual ability, and socioeconomic status. But if we are not working in a prison or juvenile hall, and if we are not concerned with identifying asocial dispositions, is there any other reason for studying this scale and for learning how to utilize it clinically? These are the questions met in the tertiary evaluation, questions which a systematic theory of test meaning must attempt to answer. We need to show, in other words, that the variable is in fact of significance in situations other than those encompassed by its domain of primary relevance.

For the So scale, three examples of such tertiary significance may be offered. The first comes from work on differential achievement among persons of unusually high intellectual talent. In 1955 a nationwide sample of high school and college students was surveyed (Gough, 1955), searching for correlates of differential academic achievement among students in the top 5% to 10% of the aptitude distribution. The one variable from the CPI which stood out above all others was the So scale: its correlations were significant for both sexes and at both educational levels. Superior achievement, as evidenced by grades, was associated with higher So scores. Note, in considering these results, that the So scale shows little or no correlation with grades if an unselected sample is studied; the CPI contains other scales, such as the two achievement indices, which are much more relevant under these circumstances.

These findings led to the theoretical formulation that consistent use of high-level talent in the academic setting is more a matter of socialization and the cathexis of approved goals and objectives than of achievement motivation per se, or of ambition as ordinarily defined and appraised, or of negative spurs such as anxiety and self-doubt.

This early study was followed by a much larger and more definitive analysis by Holland in 1959, in which freshman grades in college were forecast for a sample of 1,321 National Merit Scholarship corporation finalists. Holland found that the So scale gave valid forecasts of achievement as reflected in grades, and he also observed that the correlations for the So scale were two and three times as high as those for the Scholastic Aptitude Tests given after the finalists had already been selected.

Here, then, is one domain of functioning—academic achievement among the intellectually gifted—in which the So scale has a special significance and in which it can make a contribution which up to this time has been matched by no other variable.

A second example of a tertiary implication comes from studies of college graduation. It is known that only 40% to 50% of all students who begin college go on to graduation. This is a serious problem to educators, behavioral scientists, and everyone else. Would the So scale, administered at entry into college, be of any diagnostic value in identifying potential dropouts? The author is just finishing a large study of 3,242 students from seven classes in six colleges, in which CPI scales and Scholastic Aptitude Test scores are being compared on their forecasting efficiency. All test scores were obtained at the time of admission to college, and criteria on graduation versus dropping out were obtained 4 or more years later.

The aptitude test scores do not differeniate very well, contrary to the expectations of most psychologists. Such tests do predict GPA in college, but evidence is mounting that they bear less relationship to criteria of either rate of progress or survival (cf. Glimp & Whitla, 1964). The So scale, on the other hand, appears to be a useful predictor of graduation. Potential graduates achieve higher So scores at admission, and potential dropouts attain lower scores.

But is there an important social criterion on which persons with lower So scores excel? Certainly we should not expect or wish that in any and all situations the more conventional, more stringently self-regulated, and more interpersonally adaptable individual would do best. One such happy exception is found in the studies of creativity. Barron (1961) in his comparison of creative and journeyman writers found the former to score significantly lower on the So scale. MacKinnon (1961) in his study with Hall of 124 American architects obtained the same findings. The most creative architects scored lowest on the So scale, the least creative scored highest, and the intermediate group on creativity occupied an intermediate point on the So distribution.

CONCLUSION

What is the purpose of the method of analysis presented in this paper? Its principal goal is to specify the kind of information and comprehension needed by the practitioner of testing. The position taken is that the user of any test must become intimately and fully familiar with all of the kinds of evidence sketched here for the So scale if he is to apply the instrument in a responsible and professional manner. Analysis of the conceptual problem into primary, secondary, and tertiary components may help him to keep straight what he must do, and may indicate the kinds of information he must discover and then assimilate. Out of all this he seeks to determine what it is that he can say about the individual who has been tested. The hope is to achieve a true, profound, and individual portrait of the person being appraised.

But alas, there is no rest for the weary, and once this is done for a single scale or component of a multivariate instrument it must be done for all other components. It is a difficult task for anyone to master the meanings and implications of the variety of responses subjects make to shading on the Rorschach. But even if this mastery is achieved, is the clinician ready to interpret all other responses? No indeed, for he must come to comprehend each component in this fully dynamic and articulated fashion.

And even here we are only at the beginning and not the end of the training of the diagnostician, for once components are understood the more challenging and complex issues which arise in the interpretation of configurations and patterns must be resolved. This paper has dealt almost entirely with what one must do to arrive at an adequate comprehen-

sion of one component of one test; certainly it is evident that this is only a first step toward the kind of organized and insightful interpretation of profiles and test batteries which is the ultimate goal of the clinician in diagnosis. This inspiring and ultimate edifice of test usage, clearly, can only be constructed from building blocks of the highest quality; our need, therefore, is to delineate principles which can at each level contribute significantly to the generation of valid, dependable, and diagnostically relevant information.

REFERENCES

BARRON, F. An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology,* 1953, **17**, 327–333.

BARRON, F. Creative vision and expression in writing and painting. In, *Proceedings of the Tahoe Conference on the Creative Person.* Part II. Berkeley, Calif.: Liberal Arts Department, University Extension, University of California, 1961. Pp. 1–19.

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the mulitrait-multimethod matrix. *Psychological Bulletin,* 1959, **56**, 81–105.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychological Bulletin,* 1955, **52**, 281–302.

DONALD, E. P. Personality scale analysis of new admission to a reformatory. Unpublished master's thesis, Ohio State University, 1955.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

GLIMP, F. L., & WHITLA, D. K. Admissions and performance in college. *Harvard Alumni Bulletin,* 1964, **66**(7), 304–309.

GOUGH, H. G. A short social status inventory. *Journal of Educational Psychology,* 1949, **40**, 52–56.

GOUGH, H. G. Factors related to differential achievement among gifted persons. In P. A. Witty (Chm.), The gifted child. Symposium presented at the American Psychological Association, San Francisco, September 1955.

GOUGH, H. G. *Manual for the California Psychological Inventory.* Palo Alto: Consulting Psychologists Press, 1957.

GOUGH, H. G. The Adjective Check List as a personality assessment research technique. *Psychological Reports,* 1960, **6**, 107–122. (a)

GOUGH, H. G. Theory and measurement of socialization. *Journal of Consulting Psychology,* 1960, **24**, 23–30. (b)

GOUGH, H. G., & HEILBRUN, A. B., JR. *Manual for the Adjective Check List.* Palo Alto: Consulting Psychologists Press, 1965.

GOUGH, H. G., & PETERSON, D. R. The identification and measurement of predispositional factors in crime and delinquency. *Journal of Consulting Psychology,* 1952, **16**, 207–212.

GOUGH, H. G., & SANDHU, H. S. Validation of the CPI Socialization scale in India. *Journal of Abnormal and Social Psychology,* 1964, **68**, 544–547.

HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychological Corporation, 1943.

HOLLAND, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *Journal of Educational Psychology,* 1959, **50**, 135–142.

MACKINNON, D. W. Creativity in architects. In, *Proceedings of the Tahoe Conference on the Creative Person.* Part V. Berkeley, Calif.: Liberal Arts Department, University Extension, University of California, 1961. Pp. 1–24.

PETERSON, D. R., QUAY, H. C., & ANDERSON, A. C. Extending the construct validity of a socialization scale. *Journal of Consulting Psychology,* 1959, **23**, 182.

REED, C. F., & CUADRA, C. A. The role-taking hypothesis in delinquency. *Journal of Consulting Psychology,* 1957, **21**, 386–390.

SARBIN, T. R., ALLEN, V. L., & RUTHERFORD, E. E. Social reinforcement, socialization, and chronic delinquency. *British Journal of Social and Clinical Psychology,* 1965, **4**, 73–78.

VINCENT, C. *Unmarried mothers.* New York: Free Press of Glencoe, 1961.

WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *Journal of Consulting Psychology,* 1952, **16**, 65–72.

(Received July 27, 1964)