

THE VALIDITY OF PERSONALITY TESTS

ROBERT T. ROSS, '27

Of recent years there has appeared on the American scene a new phenomenon known as the Personality. With characteristic American zeal, we have immediately set about measuring it. Indeed we have tried to measure it before we quite know what it is or where it comes from. But this may not be a deplorable state of affairs; perhaps through continued measurement we may discover what personality is, but we must not be carried away by the appeal of statistics or our unhappy tendency to attach euphonious labels to things which we really know little about.

On the other hand, we are faced with the empirical necessity of regimenting large groups of men and women so that they will function most efficiently, and the personality test has been called upon to make its pragmatic contribution to the problem, whether or not the psychologist knows or understands just what the personality test is really doing. Many psychologists feel the same puzzled concern for this child of their laboratories as the hen who has hatched a duckling, which in turn, for reasons incomprehensible to the hen, has been able successfully to swim. The personality test has, in many instances, been able to produce results which the psychologist did not expect and does not wholly understand.

The unfortunate part of the situation lies in the fact that many psychologists and more non-psychologists, impressed with the empirical success of some of the tests, have apparently concluded that the personality test is an infallible and precise device, and on the basis of that conviction lead both themselves and their unfortunate subjects into serious psychological blind-alleys.

In the light of this state of affairs, it seems worth while briefly to consider where the personality test comes from, how it is constructed and particularly some of its limitations and misapplications.

ORIGIN OF QUESTIONS

Now the first thing to remember about a personality test, and a fact too often overlooked, is that it did not spring full-blown from the distilled essence of scientific wisdom, nor was it found one spring morning under a cabbage leaf. Some patient psychologist, somewhere, wrote down a group of questions, some of which he found in text-books, some on old notes in his pocket, some from the questions of his students, some from his own and his colleagues' experience, and some he just made up for the fun of it. This is the first form of the personality test. The next problem for the author of test was: which questions should be kept and which should be thrown out.

Of course, the choice of questions to be included in the first form of the test was not entirely haphazard. Supposing the psychologist were interested in measuring what is vaguely known as "neurotic tendency," he might include such questions as: "Do you frequently feel that you are about to 'go to pieces'?" "Do you have pronounced mood swings?", "Are you bothered by a persistent feeling of anxiety?", and the like. On the other hand, he probably would not ask, "Do you prefer to put on your left shoe first?", or "do you prefer green wall-paper to

blue?" The questions which he includes in his test in some way reflect what he considers to be the characteristics of the neurotic personality which differentiate it from the so-called normal personality. But, in general, he is at liberty to include whatever questions suit his fancy.

But a decision as to what questions he is to retain in his test and which one he is to throw out is not so arbitrary a matter. In general, there are two major methods of item selection. In the first method, the test is submitted to a group of experts, and each expert is asked to rate each question, say on an eleven-point scale, as to how indicative of neurotic tendency he considered each response to be. The psychologist can then assign to each possible response a weight which will represent the consensus concerning the degree of "neuroticism" of a person who may give the response under consideration.

THE NEUROTIC PERSONALITY

In a sense, such a test is merely an elaborate definition by experts of what constitutes the neurotic personality, and the person who takes such a test indicates by his responses how well he conforms to the definition. It may be that of the five hundred questions which the psychologist submitted to the experts for their consideration, only one hundred could be selected with sufficient consistency in the opinion of the judges to justify their use. Of these, fifty, say, are answered by the subject in such a way as to indicate neurotic tendency. We cannot, therefore, conclude that the subject is fifty per cent neurotic, but only that, *of the definitional elements included in the test*, the subject conforms to the neurotic pattern in his answers to half. It may well be that had the psychologist submitted 10,000 questions to the judges, they might have selected five hundred with sufficient consistency. And it may happen that the subject now responds "neurotically" to all of the additional four hundred items. The reliability of the diagnosis depends, therefore, upon the inclusiveness of the definition presented by the test, and it becomes important that the test include a sufficient number of items to preclude large shifts of diagnosis if the length be increased. On the other hand, it is senseless to judge the utility of a test by picking out any single item and holding it up to ridicule. The diagnosis "neurotic" depends upon the presence of many minute traits of behavior. To the extent that the subject exhibits this constellation of traits, he is presumably neurotic.

The second method of choosing items is somewhat more empirical. In this method the psychologist finds, say, five hundred persons who have been diagnosed as "neurotic" and gives them the first form of the test. He then finds a group of five hundred other persons who are as comparable to his neurotic group as possible in all traits except "neuroticism" and gives them the test. In the preceding method the expert defined the neurotic *a priori*; in this method, the neurotic is allowed to define himself, and those items are selected for retention in the test in which (statistically) a significantly different proportion of the "neurotic" group have given a certain response from that of the "normal" group. Thus if 68% of the neurotics responded

that they frequently crossed the street to avoid meeting people and only 43% of the normal group so replied, the question would be retained on the basis that a person who said he often crossed the street was more apt, in this trait at least, to belong to the neurotic than to the normal group.

TESTS ESSENTIALLY DEFINITIONAL

It should be clear then that the personality test is essentially a definitional device and that the score on the test indicates how well the subject conforms to the definition presented by the test. The great difficulty is, of course, that the constellation of behavior traits which make up any psychological characteristic is so complex that no test can hope to include them all. In medical diagnosis the problem is often much simpler. If we wished to make a test for small-pox, we could include such questions as: "Do you have a fever?", "Do you have a sore throat?", "Do you have pustules?", "Do you have a backache?", etc. No single question would make diagnosis possible, yet if the whole test were "correctly" answered, we could say with some considerable assurance that the patient had smallpox. If on the other hand, he answered "yes" to "Do you have a fever?" and "Do you have a sore throat?" and "no" to the other limited number of questions which we have included in our "smallpox" test, we could be reasonably certain that he did not have smallpox. We would not be justified, however, in saying that he was not sick. In a similar fashion, if the subject does not answer all of our "neurotic" questions in pure "neurotic" style, he may not have a "neurosis" as defined by our test, but he may still have behavioral and psychological maladjustments which interfere with his happy and efficient living.

On the other hand, it should be obvious that complete conformance to the definitional intent of the test does not guarantee the presence of the trait which the test purports to measure. In the matter of neurotic tendencies, for example, it may be that there are characteristics of the personality which remain unsampled by the test and yet so influence the personality as to offset the usual action of the traits measured, so that a person may conform to the definition of neuroticism as given by the test and still show few or none of the major maladjustments of the typical neurotic. The better the test, the greater the probability that those persons who score high on the trait measured by the test will conform to the clinical picture of such a personality; but this is only a probability and remains so by virtue of the fashion in which tests are constructed.

THE PROBABILITY FUNCTION

On the other hand, since we are dealing with a probability function, we expect to find amongst those scoring high on a test a greater proportion of persons exhibiting the trait which the test is designed to measure than amongst those scoring low on the test. The better the test, the higher the probability that the differentiation has been successful.

In the light of these facts it should be clear that it is possible for a personality test to make some contribution to our understanding of an individual or a group, and that the validity of the contribution depends as much upon the interpretation of the score as upon the score itself. If we have sufficient data concerning the construction of the test, we are forearmed and forewarned in the interpretation of results obtained by its use.

But all too frequently the psychologist who constructed the test fails to give the necessary details about the test construction while he waxes enthusiastic over his claims for his test and gives the impression that he has perfected an instrument far more valid than more objective criteria would indicate. On the other hand, personnel workers are so anxious to find a valid and reliable instrument for the measurement of personality that they are apt to let their hopes interfere with their better judgment and base conclusions on test results which are misleading and sometimes false.

SINS OF INTERPRETATION

One of the major sins of test interpretation is to assume that the test results are infallible. The most extreme example of this type of thinking of which the writer is aware was encountered in the personnel department of a large industry. In this department the practice was to use a certain personality test which purports to measure, among other variables, one called "paranoid tendency." There is considerable room for discussion as to just what the test does measure by this variable, but even its authors would probably hesitate to claim that a person having a high "paranoid" score is psychotic. On the other hand, Freud (who apparently had no use for personality tests) in another connection made the statement that it is possible that the psychotic state which is known as "paranoia" is probably the result of repressed homosexual tendencies. Putting the test and the theory together, with complete disregard for the invalidity of either or both, the personnel director whom this story concerns purported to diagnose homosexuality by the presence of a high "paranoid" score on a personality test. It is obvious what serious consequences could ensue. The writer went to the trouble to obtain some examples of high paranoid scores from unquestionably sexually normal men and some low paranoid scores from some overt homosexuals. The naivete which led to this implicit faith in the personality test is much too common in test interpretation, and although it may not lead to the absurdities and serious consequences which are implicit in the example given, it is apt to becloud the real contribution which the personality test man make with the mists of misinterpretation and misuse.

A second misunderstanding in the use of personality tests is the assumption that the test measures more than its author claims for it. This type of error is especially frequent in interest tests such as the Strong Vocational Interest Test or the Theatre Interest Test developed by Lacklen and Miller. These tests present to the subject a great variety of activities in which he may or may not be interested. With the Theatre Interest Test, for example, about 250 activities are listed including chemistry, stamp-collecting, singing, and living in the country. The test has been standardized on a group of actors and a group of non-theatrical college students. It is possible to get a score which will indicate whether the subject's interests are more in conformity with those of actors or with those of the unselected group. If, however, a subject makes a high score on the test, it is all too easy to assume that he has dramatic ability, whereas all that can legitimately be said is that he would probably find the association with theatre people pleasant. It will require much more analysis and validation of existing interest tests

(Continued on page 16)

Validity of Personality Tests

(Continued from page 8)

before they can be taken as measures of ability. The misinterpretation consists in confusing what the test measures with what it is hoped it measures. Here again it is mandatory that the person using the test understand clearly the nature of the standardization groups and the type of trait which the test measures.

A third misinterpretation is in assuming that the test measures precisely what the author claims it measures. Such a claim should not be made for a test until it has stood the test of numerous legitimate validating studies. One widely used test of neurotic tendency, for example, was so constructed that high scores were supposed to indicate neuroticism and low scores a well-adjusted personality. Validating studies showed, however, that persons scoring at both extremes of the scale tended to be maladjusted. If the intent and interpretation of the author of the scale were accepted without knowledge of the later studies, the persons using the test might very well make the possibly serious error of dismissing a behavior problem on the basis that the subject had a low score on this scale.

The necessity of being familiar with the validating studies which have been made with a personality test after it has been turned loose on the market cannot be overestimated in making legitimate interpretations of test results. It will be remembered that after Terman had constructed and standardized the Stanford Binet Test of Intelligence and had demon-

strated its ability to differentiate between groups of known intellectual characteristics, he then undertook the great validating study which is published under the title "Genetic Studies of Genius." In this study he chose 1000 children of I.Q. 140 or above, and he has now traced their development through the past 20 years. From these studies it is possible to see in actual life situations what having a high I.Q. means.

PROOF ALWAYS IN PERFORMANCE

In like manner, personality tests are constructed, standardized, and shown to be sufficiently efficient to differentiate between the standardization groups. But the proof of the test is its performance under actual testing conditions and not in differentiating between criterion groups. If a certain personality test is to be applied to the personnel of a given business or industry, it is well worth the effort to make follow-up studies of the persons testing at the extremes of the distribution of scores obtained from the test. If, for example, the neurotic tendency test referred to above had been used for the selection of personnel with the assumption that persons scoring low would be well-adjusted individuals, the test would have failed to perform as predicted by its author.

It is such conspicuous failures of personality tests to perform on a large scale that have led to the extreme criticism that the personality test is meaningless and altogether useless. This criticism is usually backed up by its proponent's quoting some single test item and asking derisively, "What can you possibly tell from that?" The answer to that particular query is given, it is hoped, in the preceding pages. As to the meaninglessness of the personality test, one can only reply that the items chosen for inclusion in the test were certainly not chosen at random, and that if their presence in the test has been justified by a standardization procedure with contrasting groups, then certainly the questions included in the test represent differing points of view of the two groups on which the standardization was based. In that sense, therefore, the test is far from meaningless. A score on the test certainly means something, and many times something important. The great problem is to define and limit exactly what the score on the test does mean. It probably does not mean precisely what its author intended it to mean, but reference to the various studies in which the test has been used will clarify its utility. Many personality tests have been proved useless, it is true, but there is no necessity for condemning them all before they are given an opportunity to demonstrate their usefulness.

A second finding also serves to lead the more impatient to discard all tests. This is the fact that in individual instances the predictions of the test fail utterly. In using an interest test for the prediction of success in the theatre, it has been found, for example, that, when the test is standardized by a group of successful actors as one group and a group of laymen as the other, if the scores of the successful actors are computed, some of them will have scores lower than those of some of the non-theatrical group. In other words, acting ability is not altogether a matter of the variables measured by the test. On the other hand, a majority of the actor group will have scores significantly in excess of those of the lay group. Such a result, as has already been explained, is to be expected on the basis



**LANE-WELLS GUN PERFORATION
"MEANS BETTER OIL WELL CONTROL"**

Lane-Wells pioneered a new technique of well control with the development of the Gun Perforator. Completion costs are lowered and the ultimate production of each well is increased. A bulletin describing the Lane-Wells Gun Perforator Service and its applications is available to engineering students. Write to Lane-Wells Company, Los Angeles, California.

LANE-WELLS



LANE-WELLS COMPANY
Technical Oil Field Services
LOS ANGELES • HOUSTON
OKLAHOMA CITY • NEW YORK

Over the Venezuelan Andes

(Continued from page 11)

elan peon, and it began to look as though we were stuck there, within a few miles of Valera, at least until one driver weakened and backed up. Within a few minutes, however, another bus pulled up behinds ours and an Army officer disembarked to locate the trouble. He ordered the bus-truck backed up some 50 yards to the nearest turnout, which the driver finally did after much grumbling; and we proceeded on our way.

We found nothing notable in Valera. A low-country town, there was much negro blood evident, and for the most part it was hot, dirty and smelly as most of Maracaibo. We stayed at a hotel which had originally been German but was now Venezuelan in all but name. The food left considerable to be desired, but at least our sleep was undisturbed.

The next morning, early, we boarded another bus and after about 30 minutes arrived in Mototan, a small village which is the inland terminus of the German-owned narrow-gauge rail-

of the probability factors which govern prediction by the use of tests.

But suppose the test is printed and distributed for use and some successful actor takes the test and fails. There is immediately raised a great hue and cry that the test is worthless. In a certain sense this criticism is true. The test is worthless if it is demanded that it measure with absolute precision in the individual case. If a subject comes to ask if he should become an actor and, having been given the test and failed it, is advised by his counselor to go into some other field, a serious error has been made. It is true that the probabilities are high that he will not be a successful actor; on the other hand, he may be one of those who may be a successful actor and still score low on the test.

The personality test, in other words, has its greatest value in dealing with the selection of personnel so that the probability of getting a group that largely conforms to the criteria implicit in the test is enhanced by the use of the test; it is, however, of doubtful value in ascribing with any high degree of certainty traits and characteristics to individuals within the group.

It is to be hoped that, as psychological techniques for the definition and measurement of human traits are perfected, the personality test will have increased utility in individual cases; until that time it should be used with care and with a full appreciation both of its limitations and of its potentialities.

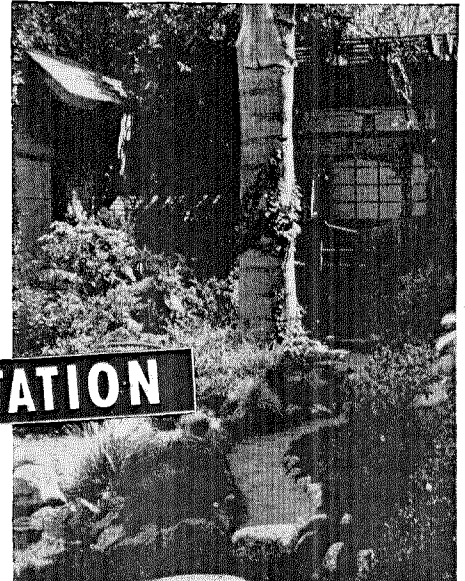
SKF puts the right bearing
in the right place."



1114 S. Hope St. PR-3168 Los Angeles

way line to La Ceiba on Lake Maracaibo. The trip was made via modern and almost new gasoline passenger car, and since we had purchased first-class tickets, we were able to sit in the front section which occupied about a third of the car's length. The rear second-class section was jammed with peons, most of them bound for one of the several villages at which we stopped en route. The trip was down hill most of the way, we did much coasting, and at every stop the car was surrounded with ragged muchachos carrying trays of "dulces" — sweets — for sale. We confined our purchases to bottled Coca Cola, which appears to be available everywhere.

The run to La Ceiba took 2½ hours. The schedule called for the twin-screw diesel passenger boat "Trujillo" to be waiting for us for the trip to Maracaibo, and our railroad tickets included the boat trip, since the boat is owned and operated by the same German company. But the "Trujillo" wasn't there, and we could not learn when it was expected. So we whiled away the time as best we could, in spite of the extreme heat, the flies, dirt and smells. No oranges were available so we again fell back on warm Coca Cola and equally warm beer. About an hour and a half later the "Trujillo" showed up — in tow. She had broken down a few miles out, and we had visions of an enforced overnight stay in La Ceiba, with accommodations of the poorest sort imaginable. However, the breakdown apparently wasn't serious, for we were able to take off about half an hour later, just before noon.



This patio is surrounded by the offices of Western Precipitation Corporation.

You are cordially invited to visit our offices and inspect our facilities for research and engineering in the field of dust, fog, fume and mist control, gas cleaning, waste recovery and the drying of solids from solutions, suspensions, slurries and filter cake. One of the most comprehensive technical libraries in the country is also at your service. Demonstrations of Cottrell Electrical Precipitators, Multiclones, Polyclones and Spray and Dispersion Dryers can be arranged for.

WESTERN PRECIPITATION CORPORATION
1016 WEST NINTH STREET, LOS ANGELES, CALIFORNIA