

CHAPTER III

PERSONALITY INVENTORIES: A RETROSPECT

3.1 INTRODUCTION

The different techniques of personality assessment were discussed in the last chapter. As the purpose of the present work is to standardize a personality inventory, detailed discussion of this technique is taken up in this chapter. It is proposed to describe how the personality inventory technique originated and developed. To illustrate this, different inventories which have been used widely, and which are types by themselves, are discussed. The description of each one of them brings out its salient characteristics. It is helpful in understanding the technique of personality inventories and its various applications. It gives a developmental picture of the technique. It would

also give an idea about the place of the proposed inventory among the members of its class.

It is very interesting to note the circumstances under which the first inventory was constructed. During the First World War a number of persons from the American Defence Forces broke down due to tension and strain when on duty at the places of action. The problem was entrusted to a team of psychologists to screen out such individuals at the time of recruitment. It was difficult to screen every individual by a psychiatric interview, and a quicker method was needed. R.S.Woodworth prepared a list of statements by consulting the psychiatrists and clinical psychologists. Every recruit was required to check the statements whether they applied to him or not. The list could be administered with brief and uniform instructions in large groups at a time. It was claimed that the method worked. This was the first inventory used to assess an individual's personality. It originated in a need for a quick and practical measure. Later on there

has been a tremendous growth of such tests. There have also been a number of researches to study their value as measures of personality. Today, more often are pointed out the limitations and the drawbacks of the method. The criticism is justified if the claim is made that it is the only and foolproof method of personality measurement. But a look back to the purpose with which it originated, reveals that till now there has been a good progress in its techniques, and it still serves the original purpose.

In the following section of this chapter, the purpose is to review the various types of personality inventories and describe a few which may characterise these types. At the end, evaluation of the inventory technique in general is discussed.

3.2 REVIEW OF A FEW WELL-KNOWN INVENTORIES

Woodworth Personal Data Sheet: The inventory referred to above, known as the Woodworth Personal Data Sheet consisted of 116 questions, to be answered "Yes" or "No". The questions, were framed on the

basis of common psychoneurotic or preneurotic symptoms about such behaviour as is found in cases of abnormal fears, obsessions and compulsions, nightmares and other sleep disturbances, excessive fatigue and other psychosomatic symptoms, feelings of unreality, motor disturbances such as twitchings, and the like¹. The "Neurotic" answer was sometimes "Yes" and sometimes "No". The total number of "Neurotic" answers was the score of an individual which was compared with the average scores of the normals and the neurotics in the standardization sample.

The item selection in Woodworth Personal Data Sheet was done on the basis of empirical criteria. The item was retained, firstly, if not more than twenty five per cent "normal" persons answered it in a "Neurotic" way; and secondly, if the psychoneurotic group consisting of previously diagnosed

1 A. Anastasi, Psychological Testing (New York: Macmillan and Co., 1961), p.494.

patients replied in an unfavourable or "Neurotic" fashion at least twice as frequently as the random group of normal people would do.

After the end of the World War I, Woodworth Personal Data Sheet was published for use with civilian population. It was also revised and adapted a number of times. It set an example in the method of construction and standardization of similar tests and also was followed in large part regarding the form and the content of the test. But for the multiple scoring procedures, many of the inventories in the current use resemble very much this ancestor.

Cornell Index: In Cornell Index is to be found the same type of questionnaire as Woodworth's. This inventory was developed during the Second World War. The items were constructed on the same basis i.e. the common psychoneurotic symptoms. The following areas of disturbances were covered in it:

Defects in adjustment expressed as feelings
of fear and inadequacy.

Pathological mood reactions, especially depression.

Nervousness and anxiety.

Neurocirculatory psychosomatic symptoms.

Pathological startle reactions.

Other psychosomatic symptoms.

Hypochondriasis and asthenia.

Gastrointestinal psychosomatic symptoms.

Excessive sensitivity and suspiciousness.

Troublesome psychopathy.

The score is compared with the performance of normal and psychiatric "rejects". An alternative scoring procedure has also been devised in which the total score is interpreted as mentioned above. And answers to particular individual items called "stop items" are given particular attention and importance in the clinical diagnosis. The examples of stop questions are:

Have you ever had a fit or convulsion?

Were you ever a patient in a mental hospital?

According to Anastasi², even the use of stop questions did not improve the validity of this instrument. It was of course designed to be only a rough screening device for personal and psychosomatic disturbances in the military selection. It is also used in civilian practice and norms for male adults are available.

In the above mentioned two examples of the inventories, the measurement has centered round the detection of diagnosis of the pathological manifestations of behaviour. In the later stages in the development of personality inventories, two distinct lines of approach have been adopted. First is a tendency to depart totally from the mere clinical diagnosis and to attempt the assessment of non-clinical traits of personality, which do not directly add to our knowledge of an individual's status regarding his adjustment or maladjustment. Such tests measure traits such as dominance-submission, introversion-extraversion, sociability, initiative, -

2 Ibid. p. 531.

responsibility and so on. Secondly, there has been a tendency to measure more than one aspect by a single test. For example, Bells Adjustment Inventory (for adults) measures adjustment or maladjustment along five areas, viz. Health, Home, Social, Emotional and Occupational. In this instance, the items belonging to the different areas are mixed together but each area has a distinct set of items. In more recent times, the items are scored for different traits. The same item is scored for a number of traits and may have same weightage in determining the score on different scales or may as well have different weightage in different cases, depending upon the scoring procedure adopted in each individual instrument. In the following pages a few examples of the non-clinical or multi-dimensional tests are described.

Allport Ascendence-Submission Reaction Study:

One of the earliest self-description inventories measuring a trait which is not a pathological manifestation but is within the normal range of personality is Allport Ascendence-Submission Reaction Study

which is still widely used. It was developed by G.W.Allport and F.H.Allport. The purpose of this inventory was to measure the tendency of a person either to lead and to dominate others or to be led and dominated by them. Each item describes a situation in which the respondent can show either dominance or submission to some degree. Following is an illustration of an item from the test:

Someone tries to push ahead of you in line. You have been waiting for some time, and can't wait much longer. Suppose the intruder is the same sex as yourself, do you usually:

- | | |
|---|-------|
| Remonstrate with the intruder | ----- |
| Call the attention of the man at the ticket window | ----- |
| "Look daggers" at the intruder or make clearly audible comments | ----- |
| Decide not to wait and go away | ----- |
| Do nothing | ----- |

Scoring weights for the A-S inventory items were determined in such a way as to differentiate persons who were rated high in dominance, by themselves and by associates, from those who were rated

low in dominance. According to Anastasi³, "considerable evidence for the validity of the total scores has been gathered, chiefly by the method of contrasted groups". This inventory in addition to its being quite popular in use, has influenced the development of many other inventories.

Bernreuter Personality Inventory: This test consists of 125 items, describing both adjustment and interests. Each item is to be answered with Yes, No or unable to answer with yes or no. Four keys were prepared on the basis of results from four previous tests: Thurstone's Personality Schedule of Neurotic Tendencies (1930), Laird's Inventory of Extraversion-Introversion (1925), Allport Ascendancy-Submission Reaction Study (1928), and Bernreuter's Test of Self-sufficiency. These four tests and the personality Inventory were administered to adults selected to represent extreme groups. Each item in the inventory was correlated with total scores on each of the four tests. The answers to each item were assigned points

3 Ibid. p.532.

on the basis of these correlations; the higher the correlation, the greater the number of points allotted. The total score was the addition of such points on all the items of a scale. These scores correlated highly with the original tests. Bernreuter's score for neurotic tendencies correlated .94 with Thurstone's schedule. Laird's and Bernreuter's introversion scores correlated .79. Allport's measure of ascendancy and Bernreuter's dominance correlated .81, and the two measures of self-sufficiency .89.

The split-half reliability of the scores on the Bernreuter's Personality Inventory was high, with median coefficient of .90. An interesting observation about these scores was very high intercorrelations between the different scales. This led Flanagan to make a factorial analysis of Bernreuter's scores of 305 eleventh-grade boys. Two factors were isolated which accounted for the intercorrelations of the four scores on the inventory. The first one was a large factor with high positive

loading on neurotic tendencies, introversion and submission and high negative loading on the self-sufficiency items. This factor was named as lack of self-confidence. The other factor, a much smaller one, was called sociability. Two new scoring keys were prepared by Flanagan to measure these two factors in addition to the four old ones. As a matter of fact these two factors are not additional scales, because, they have been derived from the old ones only. The intercorrelations between the old scales denoted that there was large overlapping in the old scales. By these two new scales, the intercorrelations are explained and overlapping is reduced, and therefore, they could be taken as substitutes for the old scales.

This inventory became very popular in use because it measured the four traits through a single administration and in less time than the original four tests. The correlations between the original tests and the corresponding scales of this inventory were also very high to justify its use. And the

scores were sufficiently stable, that is, the test-retest reliability was also high.⁴

Minnesota Multiphasic Personality Inventory

(MMPI): This is by far the most well-known of the personality inventories used in the clinical practice. It has aroused such an amount of interest among its users and research workers that a tremendously large volume of work has gone into experimentation with it. The bibliographies listed in the Third, the Fourth and the Fifth Mental Measurement Yearbooks only can give an idea about it. Moreover, it has been applied to measure more and more of personality characteristics by independent workers.

Minnesota Multiphasic Personality Inventory (MMPI) has the reputation of being a test constructed along a very systematic procedure. The authors of the inventory, Hathaway and McKinley collected about one thousand items on the basis of their own clinical

⁴ P.R.Farnsworth, "Genetic Study of the Bernreuter Personality Inventory". J.genet. Psychol. LII: 3-13, 1938.

experience, the case study records of psychiatric cases, literature on psychiatry and pre-existing personality and adjustment inventories. The items covered a wide variety of situations and subject matter and were phrased in simple language. After careful editing 504 items were retained. At a later stage more items pertaining to the Masculinity-Femininity scale were added. The scale has 550 items at present.

The items were phrased as affirmative statements in the first person. First, these statements were printed individually in large type on 3 x 5 inches cards. The testee has to divide these into three categories as "True", "False" and "Cannot say", depending upon their applicability to his own behaviour. Later on these items were printed in a booklet form and answers were to be marked on separate answer sheets.

Validation of items was carried out on the basis of criterion groups; a large group of normals (about 500) constituted the control group and for

various pathological abnormalities, groups varied from 20 to 50. Thus scales were developed for eight clinical diagnostic syndromes, a scale for masculinity-femininity and four special validating scales. A brief description of the different scales is given below:

Hypochondriasis: This scale includes worry about bodily functions. Usually the patient has a long history of exaggerating physical complaints and of seeking sympathy.

Hysteria: This scale measures conversion type symptoms, such as paralyses, contractures, gastric or intestinal complaints, or cardiac symptoms. They have attacks of weakness, fainting, or even epileptic convulsions. Hysterical cases are more immature psychologically than any other group. Although their symptoms can often be miraculously cured by a strong emotional experience, there is a great likelihood that other symptoms will appear if stress continues or recurs.

Depression: This scale measures the depths of discouragement or lack of self-confidence, which may be suicidal.

Hypomania: This scale measures overproductivity in thought and action. The patient has usually got himself into trouble because he has undertaken too many things. He is overenthusiastic and overactive, and his activities may interfere with other people through his attempts to reform social practice, or his stirring up of projects in which he soon loses interest, or his disregard of social conventions.

Psychopathic deviate: This scale measures a group of persons whose main difficulty lies in a usual absence of deep emotional responses. Nothing really matters. They are commonly likable and intelligent, but they frequently indulge in lying, stealing, alcohol and drug addiction, and sexual immorality. They may have short periods of disorientation and excitement or depression following a

discovery of their antisocial acts. They differ from some criminals in that they seem to commit crimes with little thought of possible gain to themselves or of avoiding discovery.

Paranoia: This scale shows persons characterised by suspiciousness, oversensitivity, and delusions of persecution. Patients with paranoid suspicions are common in many situations, and paranoiacs usually appear normal when on guard. They are usually quick to take vengeance against anyone who tries to control them. Persons with high scores on this scale must be handled with anticipation of this possibility.

Psychasthenia: This scale shows persons with phobias or compulsive behaviour, expressed in hand-washing, vacillation, or other ineffectual activities. The patient has queer thoughts or obsessive ideas from which he cannot escape when awake or asleep, and which serve him as a symbolic protection. Many persons, however, have phobias, such as minor fears of snakes or spiders or locked doors, without

being greatly incapacitated. As long as they can avoid these things, they get along well.

Schizophrenia: This scale measures responses which are bizarre and unusual, caused by a splitting of the subjective life of the person from reality. He reacts almost exclusively to his own thoughts, wishes, and fears. Advanced cases seldom respond consciously to the environment for long periods.

Masculinity-femininity: This scale contains items which were selected to distinguish between the two sexes in the normal group. Some items were inspired by the work of Terman and Miles.

In addition to the above mentioned nine diagnostic scales, there are four special validating scales. These are the Question score, the Lie score, the Validity score and the Correction score.

The Question Score: This consists of the total number of items classified or marked in the

"cannot say" category. The average subject places thirty or less items in this category. The more is the number in this category, automatically the scores on the other scales are lowered. The authors of the inventory say that if this number exceeds one hundred and thirty, all the scores on the diagnostic scales must be considered invalid. This interpretation, however, completely depends upon the authors' own experiences in their clinical practice.

The Lie Score: There is a tendency among the subjects to answer the questionnaires, not in a matter of fact way but in socially desirable manner. High Lie scores indicate that the subject also must have worked in the direction of lowering his actual scores on the various diagnostic scales. The Lie scale consists of fifteen items on which a completely honest person is apt to get a very low score. They are representative of socially desirable ways of behaviour which are rarely true of anyone. For example, "I always tell the truth" can rarely be answered as "yes" by anyone who is honest. Therefore,

when a subject scores high on this scale, the results are of dubious value. Here again, this interpretation is based on the authors' clinical experience and judgment.

The Validity Score: There are sixty four items which were rarely answered in the scored direction by the standardization group. The statements represent undesirable behaviour but these items together do not fall under any category of abnormality. And, therefore, it is very unlikely that an individual has all these undesirable qualities. High score on this scale indicates, according to authors, scoring errors, carelessness in responding, gross eccentricity, or deliberate malingering.

The Correction Score or K Score: This scale consists of thirty items. Twenty two items out of these are such that the subject tries to look better than what he actually is. On the remaining eight items the subject displays the opposite tendency - that of looking worse - than what he actually is.

The persons who "fake good" or "fake bad" can be detected on the basis of the composite score on this scale. This helps in differentiating between the abnormal persons whose scale scores appear to be normal and the genuinely normal and also between the normal persons whose scores appear to be abnormal and the genuinely abnormal. The low K score is indicative of self-criticizing attitude of an individual and a high score of the defensive attitude. Therefore, this score is also used as a "suppressor" score, meaning thereby, the obtained scores can be modified according to the size of this score.

There is another interesting use⁵ of this score that has been suggested by research⁶. A measure based on the difference between K score and the validity score appeared to be more useful in detecting the faking and distorting of the responses.

⁵ H. G. Gough, Simulated Patterns on the MMPI. J. abnorm, soc. Psychol., XLII: 215-225, 1947.

⁶ A. Anastasi, Psychological Testing (New York: Macmillan and Co., 1961), p. 500.

Though, these validity scales are useful in enhancing the utility of this instrument, according to Anastasi, the utilization of the various validity scales is not completely standardized, but is left partly to the judgment of the clinician.

Besides the description of this instrument there are certain other observations that also need to be considered here.

The raw scores on the MMPI scales are converted into standard score equivalents with a mean of 50 and standard deviation of 10. But the same standard score on different clinical scales does not necessarily have the same significance in the diagnosis of a case.

Various studies indicate a need for more data for the norms. The reliability coefficients obtained in different studies are on the average low for the scales of the MMPI.

MMPI can be used by the experienced clinicians only. The use of the test needs extensive

clinical experience and training particularly with the MMPI itself.

MMPI is a good instrument for general screening purposes. It also differentiates psychotics, neurotics and normals from one another. But the diagnostic validity of the separate clinical scales is questionable, especially, in the light of recent studies which employed factor analytical procedures.⁷

There has been a tendency toward interpreting scores not individually but as profiles based on all the scales. Though this method is better than that of individual scoring, here again the procedure is not standardized. A recently published "Atlas for the clinical use of the MMPI"⁸ provides coded profiles and short case histories of 968 patients, arranged according to similarity of profile pattern. This material is offered as an aid in the clinical

⁷ W.M.Wheeler, et al., The Internal Structure of the MMPI . J. consult. Psychol., XV: 134-141, 1951.

⁸ S.R.Hathaway and P.E.Meehl, An Atlas for the Clinical use of the MMPI (Minneapolis: University of Minn. Press, 1951.

interpretation of profiles.

MMPI has been used by a number of independent research workers to develop new scales. It is the greatest source of items having 550 items. Out of these the authors of the MMPI have utilized only 366 items. The rest of them are left as they are, in order not to disturb the item positions. Some of the new scales that have been designed are social introversion, dominance, responsibility, neuroticism, prejudice, and socio-economic status.

MMPI was developed for use with the abnormal cases and its application to the normal population poses problems before the users. The new scales that have been developed need further cross validation before they can be used effectively.

Guilford Factorial Inventories: In the inventories so far discussed the factorial studies were only incidental and were conducted after the construction and standardization of inventories was over. In this case, the construction of the inventory was

based upon the factor analysis in order to avoid the overlapping among the different factors that were measured. According to its authors several factor analyses served as starting point. In the first analysis, the hypothesis that the dimension of introversion-extraversion represented an independent personality factor or primary trait was put to test.⁹ Similar studies were conducted to investigate other areas of temperament.¹⁰

When originally constructed there were three inventories. The first one of these, "An Inventory of Factors STDCR" measured five factors, viz:

S Social introversion; seclusiveness: A high score indicated sociability, a tendency to

9 J.P.Guilford and R.B.Guilford, An Analysis of the Factors in a Typical Test of Introversion-Extraversion . J.abnorm. soc. Psychol. XXVIII: 377-399, 1934.

Supra., Personality Factors S,E and M and Their Measurement . J.Psychol. II: 109-127, 1936.

Supra., Personality Factors D,R,T and A . J.abnorm.soc.Psychol. XXXIV: 21-26, 1939.

10 Supra., Personality Factors N and GD . J. abnorm. soc. Psychol., XXXIV: 239-248, 1939.

seek social contacts and to enjoy the company of others. A low score indicated shyness, a tendency to withdraw from social situations.

T Thinking introversion; reflectiveness: A high score indicated a lack of introspectiveness and extravertive orientation of the thinking processes. A low score indicated an inclination to meditative thinking, philosophising, analysing one's self and others, and an introspective disposition.

D Depression; unhappiness; pessimism: A high score indicated freedom from depression, a cheerful, optimistic disposition. A low score indicated a chronically depressed mood including feelings of unworthiness and guilt.

C Cycloid disposition; emotional instability: A high score indicated stable emotional reactions and moods and freedom from cycloid tendencies. A low score meant strong emotional reactions, fluctuations in mood,

and a disposition towards instability.

- R Rhathymia; carefreeness; happy-go-lucky disposition: A high score meant carefree disposition, liveliness and impulsiveness. A low score meant an inhibited disposition and an overcontrol of the impulses.

The second inventory "The Guilford-Martin Personnel Inventory II" measured the following three factors:

- O Objectivity vs. subjectivity or hypersensitivity: A high score on this indicated a tendency to view one's self and surroundings objectively and dispassionately. A low score meant a tendency to take everything personally and subjectively and to be oversensitive.

- Ag Agreeableness vs. generalized hostility; belligerence: A high score meant a lack of quarrelsomeness and a lack of domineering qualities or tendencies. A low score indicated a belligerent, domineering attitude and

an overreadiness to fight over trifles.

Co Cooperativeness (or tolerance) vs. fault finding disposition: A high score indicated an overcriticalness of people and things and intolerance.

The third inventory "The Inventory of the Factors GAMIN" contained the undermentioned five scales:

G General drive for activity; energy: A high score indicated a tendency to engage in vigorous overt actions. A low score indicated inertness or lack of interest in motor activity.

A Ascendancy (social boldness) vs. submission: A high score indicated leadership qualities in social situations and low score, passive or submissive tendencies.

M Masculinity (of emotions and interests) vs. femininity: A high score indicated manliness

of emotional and temperamental traits and low scores, femininity of these.

I Inferiority feelings vs. confidence: A high score indicated self-confidence. A low score indicated lack of it, underevaluation of one's self, and feelings of inadequacy and inferiority.

N Nervousness vs. composure; calmness: A high score indicated a tendency to be calm and relaxed; a low score indicated jumpiness, a tendency to be easily distracted, irritated or annoyed.

In the first inventory, the factors S, T and R seemed to measure the different aspects of the same common trait introversion-extraversion. The remaining two factors also seemed to fall under another common concept of neuroticism. In the third inventory, the factors of Inferiority and Nervousness belonged again under the above mentioned neuroticism trait. Therefore, it was decided to combine the

three into one single inventory, and to avoid high intercorrelations found among the original factors. The new inventory known as the Guilford-Zimmerman Temperament Survey (GZTS) measures the following traits:

- G General activity
R Restraint vs. Rhythymia
A Ascendance
S Sociability
E Emotional stability (combination of D and G)
O Objectivity
F Friendliness (previously called "agreeable-
ness")
T Thoughtfulness (previously, T)
P Personal relations (previously, Co)
M Masculinity (of emotions and interests)

Just as in the MMPI, there are three validation scoring keys for detecting (i) those who willfully try to fake good, (ii) those who do it without realizing it themselves, i.e. unconsciously, and (iii) those who are careless and erratic in -

responding.

Guilford Inventories were a step toward progress, especially, due to the method of factor analysis employed in the construction of the different scales. But it was found that still there were appreciable intercorrelations among the different scales of the original three inventories. Even though, in the GZTS composite inventory, only ten traits were retained, Thurstone¹¹ analysed the Guilford data to conclude that only seven major factors were enough to account for the intercorrelations instead of the original thirteen.

The reliability coefficients for different scales in the original inventories ranged from .80 to .94 and for those in the GZTS from .75 to .87. The evidence for the empirical validity of the different scales is insufficient. There are a few -

¹¹ L. L. Thurstone, "The Dimensions of Temperament". *Psychometrika*, XVI: 11-20, 1951.

incidental studies reported by Guilford¹² in this connection but they are by no means conclusive.

Gordon Personal Profile: A new type in inventory construction was introduced with the application of a forced-choice technique. Though this technique is discussed in its details separately, the Gordon Personal Profile as one example of an inventory of this sort is discussed here. This inventory comprises of seventy-two descriptive statements grouped into units of four each, called tetrads. The inventory measures four traits and each tetrad contains an items belonging to each one of these traits. The factors that are measured are:

- | | |
|---|---------------------|
| A | Ascendancy |
| R | Responsibility |
| E | Emotional stability |
| S | Sociability |

The definitions of these scales are given

¹² J. P. Guilford, *Personality* (New York: McGraw Hill Book Co., 1959), pp. 185-187.

below from the Manual¹³ of the test.

Ascendancy: Those individuals who adopt an active role in group situations, who are self-assured and assertive in relationships with others, and who tend to make independent decisions, make high scores on this scale. Those who play a passive role in the group, who would rather observe than participate, who generally lack self-confidence, who prefer to have others take the lead, and who tend to be overly dependent on others for advice, normally make low scores on this scale.

Responsibility: Those individuals who take responsibilities seriously, who are able to stick to any job and get it done, who are persevering and determined, score high on this scale. Individuals who are unable to stick to tasks that do not interest them, and in the extreme, who tend to be flighty or irresponsible, usually make low scores on this

¹³ L. V. Gordon; Gordon Personal Profile, Manual (New York: World Book Co., 1953).

scale.

Emotional Stability: High scores on this scale characterise individuals who are well balanced, emotionally stable, and relatively free from anxiety and nervous tension. Low scores are associated with excessive anxiety, tension, hypersensitivity, and nervousness. Large negative scores may indicate the traditional "neurotic".

Sociability: High 'S' scores are made by individuals who like to be with and work with people, who are gregarious and sociable. Low scores reflect a lack of gregariousness, restriction in social contacts, and in the extreme, an avoidance of social relationships.

There is also a measure of total score called 'T' which does not have much meaning by itself. It is only an algebraic sum of the favourable and unfavourable responses.

The test is applicable to the adolescents

and the adults.

Each tetrad contains two favourable statements equated for high preference values* and two unfavourable statements equated for low preference values. The subject is required to select one item that is most applicable to him and one that is least applicable to him.

Reliability coefficients calculated by various methods range from .74 to .95 with majority of them around .85. The data reported on the validity of the test are also quite promising and show a relatively higher validity of the different scales of the Personal Profile. Even though the scales were developed on the basis of factor analysis, there is still some amount of correlation especially between Ascendancy and Sociability and between Emotional Stability and Responsibility.¹⁴

* The preference value of an item is a measure of the extent to which people in general tend to rate themselves "high" or "low" on the characteristics described in that item.

¹⁴ A. Anastasi, Psychological Testing (New York: Macmillan and Co., 1954), p. 546.

The different types of personality inventories have been described and evaluated so far individually. In general, inventories as a whole are criticised very strongly. In the following section attempt is made to discuss briefly the major points of such criticism.

3.3 GENERAL EVALUATION OF THE PERSONALITY INVENTORIES

There is tremendous amount of growth and use of the personality inventories. This is, in spite of the fact that they are so vehemently criticised by extremists who would like to ban this instrument from the field of psychological testing. Most of the psychologists, however, occupy the intermediate position where they make use of the inventories with the full awareness of their limitations. It is, therefore, necessary to see what are the major points of criticism against the use of the personality inventories in general.

1. The behaviour of an individual is more changeable in the areas covered by the personality

tests than those covered by ability and aptitude tests. In other words the personality variables are not as stable as the ability variables. But this fact does not disqualify personality inventories in particular. Rather it is one of the unavoidable obstacles in the field of personality measurement, whatever be the method used. It, particularly, poses a problem for determining the reliability of the personality tests as such. When the behaviour itself is subject to change, the inconsistency of responses cannot be solely attributed either to the method of assessment or the behaviour itself. But the need for such assessment is so great that one has to tolerate this fact on the grounds that the deeper and more subtle patterns of behaviour are more enduring and once formed, do not change so easily. This makes it possible to measure and predict behaviour with sufficient accuracy.

2. Sometimes one does not know whether the behaviour itself has undergone change, but the

responses are changed. Guilford¹⁵ has reported studies in this connection. Eventhough there are some changes, they often act in the opposite directions to neutralise each other and the total effect due to such changes is negligible.

3. An individual does not behave consistently in all the situations. For example, one who is extravert and sociable in a classroom may not behave in the same manner at home and among relatives. If the items in an inventory cover some narrow field of behaviour, then, of course, this point of criticism stands. But as a general rule in the item construction, the area of behaviour should be covered as widely and thoroughly as possible. If this is not done it is a drawback of that particular test and not of the personality inventories in general.

4. The meanings attached to the trait-names by different persons are different in many cases and may widely differ from one attached by the authors

15 J.P.Guilford, 1959, Op.Cit. p.193.

originally. This happens almost invariably when the authors do not define clearly the different terms they use in their manuals. There is not a standard terminology and different authors name the traits different^{-ly} and sometimes in quite novel ways. The users on the other hand go merely by the commonsense usage of the terms instead of going through the manual for more exact meaning. This can happen in any kind of personality tests and not merely in the self-report inventories. This is moreover a difficulty in the procedure of test administration and interpretation and can be overcome by careful definitions and accurate interpretations.

5. Some critics say that the examinee does not know himself well enough to make a dependable self-report. Even though it sounds absurd to some, there is still some element of truth in it. So far as the questions asked are simple, such as, "do you get nightmares" or "do you keep a diary", it is very easy to report correctly. Majority of the items are of this kind, or rather they should be of such kind. But

sometimes the subject comes across an item which does not merely ask to report a fact. It requires him to give his judgment or interpretation of a situation, which might well be beyond his ability. Firstly, such cases are rare. It is never so difficult. Anyone can reasonably be expected to think for himself, make judgments and interpret facts. Secondly, during item analysis, the items which cannot be understood properly, which are ambiguous, and which are beyond the grasp of the group on which the test is being standardized or is standardized, are most likely to be eliminated. Even though some such items remain, Guilford¹⁶ says that whatever the subject reports is significant for him. Only thing is the response should be properly keyed. Even if the subject misrepresents the facts, it can be taken as a significant indicator of his behaviour mechanisms.

6. Different examinees interpret an item in different ways. Even though the examinees are

16 J.P.Guilford, 1959, Op.Cit. pp.191-192.

generally supposed to have a common cultural and educational background with those included in the standardization sample, and even though the bad items are dropped through item analysis, there is bound to be some possibility that a few items can be misinterpreted or rather interpreted differently by different persons. Firstly, one or two of such items would not matter if the rest of the items are carefully constructed and edited and have gone through rigorous item-analysis procedures. There is no need to despise the value of the entire inventory if one or two bad items can be detected on this ground. Secondly, Guilford¹⁷ puts forth a different point of view. According to him if an item has gone through the item-analysis process, it has some validity. It might be even due to the fact that the item is ambiguous and is differently interpreted. It becomes a kind of projective test. "If an item predicts or indicates trait positions of individuals, it does so in spite of, and perhaps in some cases because

17 J.P.Guilford, 1959, *Op.cit.* pp.193-194.

of, misinterpretations."¹⁸

7. Examinees are not always honest in answering the personality tests. The question arises that, "do all the examinees invariably falsify their answers?" The fact that there is always a scope for malingering in the self-report inventories, does not imply this. It can be done but it is not done invariably. For example, an applicant for a job would try to appear good by answering in the more desirable direction; but one who comes for the solution of his difficulties to a counselling centre has more reason to be honest in his responses. During the Second World War, the prospective recruits tried to fake bad because they knew that emotionally unstable individuals were not sent for dangerous and taxing situations and under conditions of stress. So the malingering depends upon the purpose of taking the test. Edwards¹⁹ conducted an investigation

¹⁸ J.P.Guilford, 1959, Op.Cit. p.194.

¹⁹ A.L.Edwards, "The Relationship between the Judged Desirability of a Trait and the Probability that the Trait will be Endorsed"., J.appl. Psychol., XXXVII: 90-93, 1953.

to determine the amount of relationship between the social desirability of a trait and the probability of the trait being endorsed. The relationship found was very high ($r = .87$). Hanley,²⁰ later on, in an independent enquiry, confirmed this relationship. But there are a number of ways in which it can be interpreted. Firstly, the examinees bias their answers in the direction of social desirability. Secondly, the socially desirable qualities are more common among people. Thirdly, what qualities individuals have, they consider those to be desirable. In the experiments,²¹ the possibility of influencing the test scores by giving instructions to do so was studied and it was found that it was quite possible. But this does not necessitate a total ban on the use of inventories. It is necessary to be more cautious. In fact, as seen in the case of MMPI and the Guilford-Zimmerman inventories, there are some devices such

20 C.Hanley, "Social Desirability and Responses to Items from Three MMPI Scales"; D,Sc,and K".J.appl. Psychol., XL: 324-28, 1956.

21 R.G.Bernreuter, "The Theory and Construction of Personality Inventory". J.soc.Psychol., IV: 387-405, 1933.

as validation scores which act as checks against such practices. More recently the forced-choice technique has come to be explored as a very effective check against this.

8. According to Nunnally²² "self-description inventories are usually less reliable than tests of aptitude, achievement, interests and attitude". He also admits that there are numerous exceptions to this statement. The reliability of personality inventories is first of all affected by some of the factors we have already considered. The personality variables themselves are more changeable than the ability variables. There are a number of other factors too. It is better if we consider the reliability of the inventories with other methods of personality assessment. It is a major point ~~that~~ in favour of their use, because, the other methods, especially the most common in use such as rating

²² J. C. Nunnally, Jr., Tests and Measurements; Assessment and Prediction (New York: McGraw Hill Book Co., 1959), p.331.

scales and projective techniques are the most notorious for their low reliability. With proper care in the item construction and by increasing the number of good items, the reliability of a personality inventory can be raised easily to a desirable level. Most of the inventories have it between .75 and .85. A few have even higher than .85.

9. Another criticism against inventories is that majority of them lack empirical validity. They depend on the most part on the face validity or content validity. It is proper to start with the content validity but ultimately the data for the empirical validity should be gathered to justify its use. It is many times a practice to correlate one inventory scores with those of another. But this is not a clear indication of test validity. The validity also differs from one situation to another. As has already been said, the applicants for a job enhance their scores in the desirable direction and if these scores are correlated with their future job success, the relationship might be very low.

Because of this factor they have very little use in the selection programmes. The results in the field of predicting success in the academic field are varied. The correlations range from .08 to .44 in different cases and with different populations.²³ In the field of pathology, the inventories differentiate between the normals and the broad categories of pathology such as general neurotics or psychotics. But their validity in diagnosing the more specific clinical disorders is still questionable. In the field of predicting vocational success the validity differs from occupation to occupation.²⁴

10. Another point of criticism is the response set or response bias in answering. A subject might be more prone to answer "yes" rather than "no" or vice versa. This certainly vitiates the actual score in either negative or the positive direction, depending upon the nature of the items. This is not

23 J.P.Guilford, 1959, Op.Cit., p.201.

24 E.E.Ghiselli and R.P.Barthol, "The Validity of Personality Inventories in the Selection of Employees", J.appl.Psychol., XXXVII: 18-20, 1953.

a criticism of the personality inventories in particular because this applies to tests of abilities, aptitudes and achievement as well. Generally, the correct or keyed responses should be distributed evenly among the different alternative positions of the answers. The validation scores detect carelessness in responding or such other factors and act as check against these.

Though these are the various criticisms against the personality inventories, they do not warrant ban on their use. There are explanations and ways out. There are no other tools that can be so easily handled by moderately trained workers in the field of psychological testing. The need for the assessment of personality qualities is so great that the few experts who can profitably use the projective or other methods cannot be solely depended upon. Moreover, the value of other methods is also greatly debated and questioned. In the present circumstances, therefore, any attempt to improve upon the technique which can be most widely used is

welcome and needs due encouragement. In Nunnally's words, "the great need to measure personality characteristics and the paucity of adequate measures should make us cautious about disparaging any well-intentioned efforts."²⁵

3.4 SUMMARY

Personality inventories originated in the First World War as a quick screening device. Since then there has been a tremendous growth in their number and the variety. Different inventories serve different purposes. Some are suitable in clinical use, some in counselling, and some in vocational guidance or selection. The content of each inventory depends upon its purpose. Some inventories measure single factors, while others measure more than one. All of them are based on the principle of self-rating.

Even though they are most widely used, there is vehement criticism against them. If their role is understood properly as a technique which is easy,

25 J.C.Nunnally, 1959, Op.Cit., p.336.

quick, reliable and economical, they serve the purpose very well. The points of criticism often lead to the development and improvement of the technique. In this case also, the criticism about faking behaviour of the subjects led to the development of the lie scales and the forced-choice technique. The present inventory is based upon the forced-choice technique, which is discussed in the next chapter. The general procedure of the standardization of the present inventory is also given in the next chapter.

REFERENCES

1. Anastasi, A., Psychological Testing. New York: Macmillan and Co., 1954 and 1961.
2. Bernreuter, R.G., "The Theory and Construction of Personality Inventory". J.soc.Psychol., IV: 387-405, 1933.
3. Edwards, A.L., "The Relationship between the Judged Desirability of Trait and the Probability that the Trait will be Endorsed". J.appl. Psychol., XXXVII: 90-93, 1953.
4. Farnsworth, P.R., "Genetic Study of the Bernreuter Personality Inventory". J.genet. Psychol., LII: 3-13, 1938.
5. Ferguson, L.W., Personality Measurement. New York: McGraw Hill Book Co., 1952.
6. Ghiselli, E.E. and Barthol, "The Validity of Personality Inventories in the Selection of Employees". J.appl. Psychol., XXXVII: 18-20, 1953.
7. Gordon, L.V., Manual: Gordon Personal Profile. New York: World Book Co., 1953.
8. Gough, H.G., "Simulated Patterns on the MMPI". J.abnorm. soc. Psychol., XLII: 215-225, 1947.
9. Guilford, J.P., Personality. New York: McGraw Hill Book Co., 1959.
10. Guilford, J.P. and R.B.Guilford, "An Analysis of the Factors in a Typical Test of Introversion-extraversion". J.abnorm. soc. Psychol., XXVIII: 377-399, 1934.

11. ----- "Personality Factors S, E and M and Their Measurement". J. Psychol., II: 109-127, 1936.
12. ----- "Personality Factors D, R, T and A". J. abnorm. soc. Psychol., XXXIV: 21-36, 1939.
13. ----- "Personality Factors N and GD". J.abnorm. soc. Psychol., XXXIV: 239-248, 1939.
14. Hanley, C., "Social Desirability and Responses to Items from Three MMPI Scales: D, Sc, and K". J.appl. Psychol., XL: 324-28, 1956.
15. Hathaway, S.R. and P.E.Meehl, An Atlas for the Clinical Use of the MMPI. Minneapolis: Univer. Minn. Press, 1951.
16. Nunnally, J.C., Jr., Tests and Measurements: Assessment and Prediction. New York: McGraw Hill Book Co., 1959.
17. Thurstone, L.L., "The Dimensions of Temperament". Psychometrika, XVI: 11-20, 1951.
18. Wheeler, W.M., K.B.Little and G.F.I. Lehner, "The Internal Structure of the MMPI". J. consult. Psychol., XV: 134-141, 1951.