

CHAPTER III

GENERAL PROBLEMS IN PERSONALITY MEASUREMENT

A critical weakness in the scientific study of personality is the relative lack of adequate measurement operations. While many tests, instruments, and assessment devices are available, each has its basic limitations. It is often held that the more penetrating techniques are not objective, while the so called objective instruments are superficial. In the personality area, tests and other measurement devices are the only instruments by which systematic data are collected. It is axiomatic that measurement is necessary in the scientific study of personality. Both placing a person in a diagnostic category and preparing a case study are measurement operations, broadly constructed. Of more general

significance is the role of measurement in testing the theoretical propositions embedded in a given systematic view of personality.

Measurement operations can also serve the important function of providing data permitting the refinement or revision of concepts. If the indices for two concepts covary highly among people, the investigator must consider the possibility that one new concept should replace the pair.

Measurement requires no theory. When a person sets out to measure things, he needs to have in mind only a variable or a classification scheme. Measurement is theoretical but should not be anti-theoretical. The few assumptions on which it is based should be acceptable to most personality theorists. These are: (a) Science in general, and personality in particular, is a theoretical construction that exists in the minds of people, (b) There is some consistency over time in the manner or mode of each person's interaction with his environment, (c) There is sufficient regularity and similarity in the overt behaviour of different individuals to permit the fruitful description of these individuals in terms of a common set of variables. In this connection, a few problems relevant to personality measurement should be examined and understood adequately. These are stated below.

Problem 1 : The Inadequacy of Definitions:

Most of the variables in personality theories have definitions that are insufficient from the viewpoint of measurement. The limitations stem primarily from the nature of personality: its variables are not as intuitively obvious as some of the basic definitions of physics nor as tied to specific operations as other dimensions in the natural sciences. Moreover, the observations generating personality concepts are made in the clinic or in the unsystematic world of everyday experience, places in which it is difficult to specify the antecedent conditions with any precision.

An indirect consequence of the breadth of the conceptualization of the typical personality variable is that its definition is not sufficiently explicit and specific. The delineation usually fails to indicate the stimulus condition in conjunction with which the dimension may be assessed. While this unfortunate omission may stem from an assumption that these conditions are self-evident, it tends to suggest that the appearance of the behaviour coded by the variable is spontaneous or internally generated regardless of the external setting. In addition, it unfortunately permits the experimenter to use any of a wide variety of conditions for his assessment, and thus greatly

decreases the possibility of comparability of measurements in different research studies.

A second kind of personality variable is that which is identified by a technical term not defined in the abridged dictionary, a word coined for a special purpose. Examples are "anality", "incongruence", or "endocathection". Variables designated in this way have one important advantage over those of the first kind. While they are less readily comprehended, they have less surplus meaning derived from literary and loose common parlance.

In order to solve these problems, first step must be to prepare a complete statement of the variable. Such an explication may well include a word or phrase for economical identification of the variable. It should then go on to specify the responses subsumed under the variable and the one or more sets of conditions under which these responses may be observed for purposes of measurement.

The specification should include not only the environmental setting but also the internal state of the person. The description should also bring out the basis for estimating the strength of the disposition.

In delineating a personality variable, the investigator must make explicit both the high and low

end of his continuum. A continuum having compulsive concern at one end should extend through appropriate emphasis on both details and overall considerations to preoccupation with wholes at the other.

This conceptual problem should be differentiated from the operational aspect. Many scales are designed to measure the extent of a person's tendency to show a particular kind of behaviour which in the extreme is maladaptive.

When an investigator has such a delineations of his variable, he will be able to judge the degree of congruence between his concept and any actual or planned set of operations for measuring the variable. He can determine not only the extent to which a single test or instrument embraces the domain of the concept but also the probability that the procedure will produce measurement with variance from irrelevant sources.

There are some personality variables that do not fall into either of the two classes. These are variables defined in terms to specific operations: For example, movement as noted in Rorschach responses manifest anxiety as measured by Taylor scale, and leveling as determined by the Squares Test. The investigator using such variables can immediately turn to two types of questions: what are the

psychometric properties of the variable? What is the theoretical significance of the concept? The first question can be answered in usual ways, for example, by determining the intercorrelations of the part scores and by determining the stability of the index over time. The second question can be studied by empirical investigations of the relationships between this variable and others with which it has apparent similarity or postulated theoretical connections. The problem of inadequacy of definition could be completely avoided by restricting experimental work to promising variables of this kind. Such an approach to the scientific study of personality may well advance the field more than persistent efforts to subject to empirical test theoretical propositions about concepts so broadly defined that any one set of measurement operations can reflect only a tiny portion of their domains.

Problem 2 : The Probabilistic Nature of Response Tendencies:

Personality is concerned with tendencies, with dispositions, rather than with such all-or-none, present-or-absent matters. More exactly, the strength of a personality variable in a given person is indexed by the proportion of times that he makes a particular response in specified situations. In any situation, a person can be observed to make different responses at different times, he has many

potential responses available to him. The responses observed on any one occasion are only a sample of potential responses. This problem has qualitative and quantitative aspects. In a relatively unstructured situation, responses indicative of discrete variables compete. In principle, the two persons may have the same strength of one tendency but have different proportions of responses manifesting the disposition because of differences in the strengths of one or more other variables.

For more fruitful comparisons of persons with respect to a single variable, the situation must be structured so that the observed responses can be scored to provide an estimate of that variable. It has been observed that even under such a condition, much variability of responses repeated on a second testing after an interval of one or more weeks vary from .16 to .80. There is the additional complication that, for each separate procedure, reliable individual differences are usually found in the extent of variability in responses over time. Such differences appear to be associated with the specific instrument, or at least with a type of instrument, rather than themselves indicating a general characteristic of the individual.

In both unstructured and structured testing situations, the rare response is less stable than the common one. Idiosyncratic responses are less likely to be repeated than popular ones. The occurrence of an usual response can be used as evidence for the presence of a disposition, although the absence of that response does not necessarily indicate a low strength for that tendency. However, an unusual response can occur in individuals, varying widely in strength of disposition, the observation only indicating that the strength is above some minimal level.

In structured procedures presenting the appropriate setting for assessing a given variable, the variability problem takes the form of instability of obtained measurements. The scores for any one day may have high homogeneity but have an unsatisfactorily low correlation with scores obtained on another day. The set of responses given by a subject on one day is influenced by factors affecting that set of responses while other factors affect the set given on the second testing. Here the extraneous variance, while systematic, is not readily attributed to other personality variables; rather it is due to the lack of experimental control over the subjects' experiences prior to the testing. Such factors obviously reduce the correspondence between the measurements and the concept.

In order to cope with the problem of the probabilistic nature of response tendencies, two possibilities are available. The preferred approach is to look for sets of stimuli and for testing conditions that will yield scores of appropriate stability. Everyday observation indicates that personality variables should not be construed as highly stable dispositions, unvarying over time. While a case could be made for maximizing the stability of measurements pertinent to a variable so that test stability was considerably higher than the postulated stability of the variable, such an effort would seem contain to reduce fluctuating variables such as mood.

A second mode of coping with the problem is expensive and therefore less practical. One could administer a test on each of the several occasions, to minimize the contributions of bias associated with any single day. In the abstract, this is the ideal approach; however, its drawbacks are several. To gain access to subjects on several occasions not only is costly and difficult but also the effects on responses of repeated exposures to the test materials must be considered. This orientation therefore deserves the expenditure of considerable time and effort to determine its potential value.

Problem 3 : The specificity of Personality Measurement:

Most if not all procedures for measuring personality yield scores containing a high proportion of method variance. Analytically, this variance can have several sources. An obvious and not unexpected source is the person providing the data. A subjects' description of himself will understandably depart from his description as provided by other people. In addition, persons in different relationships to the subject may see him differently. Method factors may also be associated with type of item content, with format of items, used with test instructions. Finally there is variance associated with the total setting, with the meaning of the situation for the subject. The same subjects taking a test anonymously for research purposes may have a quite different distribution of scores than would be obtained if the test were part of the requirements for admission to a school or for employment.

The significance of this specificity associated with individual tests is probably underestimated by most researchers in the personality field. Yet, any one surveying the literature on any given variable cannot fail to be impressed by the fact that independent methods of measurement show little, if any, association. The investigator should make

every possible effort to determine the kinds and extent of method factors in his measuring instruments. He should also avoid the economical but dangerous practice of restricting himself to a single instrument, but rather should employ a minimum of two procedures as dissimilar in method as possible. Only when he can demonstrate comparable findings from different ways of measuring his variable, can he begin to hope that he is getting at the core of his concept and is not misinterpreting systematic method variance as trait variance.

This aspect of measurement has become the main problem of study in the present investigation.

Problem 4 : The Distortion resulting from Individuality:

Measurements of different individuals cannot meaningfully be compared unless the conditions have been identical for the several applications of the procedure. Two persons with the same strength of a behavioural tendency may obtain different scores if the test for that disposition is given under different conditions. In measuring personality, it is usually assumed that the stimuli and the situation have the same meaning for all subjects, so that any differences in scores are functions of real differences on the trait being measured. This assumption is questionable. Individual differences in the meaning of the test situation are another source of possible distortion of measurements.

Finally, there is the theoretical problem of individuality in the manifestation of a trait within different persons and in the role played by that trait within diverse personality structure. Two exceptions may be noted for the problem of distortion due to individuality. In the first place, techniques involving relatively unstructured situations and ambiguous stimuli capitalize on the subject's idiosyncratic interpretations and perceptions. They obviously require a highly skilled and perceptive examiner who can infer the subject's view of the test situation and can also detect specific influences unique to the given conditions, such as the effects of his own feelings toward the subject. In the second place, the problem of individual interpretation of test items is not present in tests developed by empirical keying. In such an approach, the meaning of the item is not assumed a priori, but rather resides in obtained correlations with independent criteria.

The investigator must face this problem of individuality in his development of the stimuli and the setting for measuring a personality variable. In most instances, he should seek manifest evidence that the situation has the same significance for all subjects and that the subjects are homogenous in responding to the items.

Problem 5 : The Peculiar Situation in which Personality
is Measured:

Measurement of different objects can be compared only if the same procedures were employed in making the observations so that the measurements are determined solely by the different degrees of the measured quality possessed by the various objects. Such conditions are ordinarily met in tests of ability. The instructions to the subject as well as the nature of the items make clear that ability is being assessed. Subjects know what is meant by the notion of a "right answer" and they try to give as many as they can. The examiner attempts to structure the situation so that motivation will be optimal. He also seeks to prevent the mobilization of anxiety, emotion and excessive drive that could impair performance. Furthermore, it is apparent to the subject that the examiner is working with him to maximize the score. As a consequence, in ability testing, there is sufficient consensus among subjects on the meaning of items, the task and the situations so that comparisons between scores can be made with reasonable confidence.

Personality testing differs from ability testing in almost everyone of these respects. The instructions do not specify that personality is being measured but rather use some innocuous phrase such as "What you are like". The

subjects are told that there are no right answers, but that they should indicate what is true for them; the criteria for determining what response they should make are certainly vague. The examiner wants to find out something of which the subject is probably unaware, a condition that understandably generates anxiety and possibly antagonism in the subject. These differences between the conditions for testing ability and personality undoubtedly contribute a great deal to the lower stability, lower internal consistency and lower correlations with other variables that are generally found for measures of personality.

One way to meet this problem is to make the most of the lack of structure by determining what motivations emerge when subjects confront an ambiguous situation, classifying subjects according to the needs manifested, and building a science around the resulting typology. A better way to cope with the difficulty would seem to be to develop procedures that will elicit homogeneous interpretations by subjects but will still provide indices of the strengths of dispositions. Such procedures may set a reasonable task that the subject can willingly undertake, a task which, however, will involve responses that differentiate persons on relevant variables.

Problem 6 : The Question of Representativeness:

The objective of personality measurement can be stated as the description of the subject as he usually behaves, not as he can be expected to respond in the artificial and temporary conditions of the assessment situation. In contrast, the purpose of testing ability is not such an appraisal of normal functioning, but rather the estimation of potential ability, of maximum capacity under optimal condition. Perhaps one should attempt to assess a person's maximum tendency toward a particular kind of behaviour under conditions designed to provoke it or to facilitate its appearance. This approach could be feasible for socially desirable traits such as co-operativeness and possibly also for leadership or initiative. In principle, undesirable characteristics might be measured in meaningful, seemingly naturalistic situations that imposed stress on the subjects.

In addition to the atypicality of the laboratory or testing situation, the subject is generally unfamiliar with the instructions, the stimuli and the responses. Perhaps of greatest importance is the limitation placed on the subjects' responses. In ordinary situations, a person has considerable or complete freedom with respect to the form or structure of his behaviour, and much freedom as to its content. In tests,

the subject is typically restricted to choosing one response from among a limited set of alternatives, all of which may seem to be categorical, inappropriate, or unnatural for him. Thus, it may not be only the unrepresentativeness of the situation and stimuli but also atypicality of the permitted responses that reduce the possible generalization from personality measurements.

A testing instrument in which each separate item is keyed for any one variable is unrepresentative in still another respect. In most everyday situations, one tendency is more or less in competition with other tendencies; one motive may be in conflict with another; but we tend to measure only one variable at a time. This approach is probably a sound initial procedure.

This problem can be solved not by armchair speculation but by empirical study. By analogy, it is suggested that, in principle, it should be possible to make measurements in a testing room, that can be generalized to everyday behaviour. The empirical evidence for the specificity of personality measurements, however, provides little basis for optimism about success in this endeavour. On the other hand personality variables with narrow definitions specifying the conditions under which the tendencies may be observed may well prove amenable to such generalizations from the laboratory or testing room.

Closer scrutiny of the problem suggests that, in operational terms, it is meaningless to talk of the way a person usually behaves. The best way that can be done is to have an expert judge to process a large and representative body of observation protocols concerning a subject and reduce them to a single index of central tendency through an implicit weighting of the diverse data. Alternatively, a series of separate measures can be objectively combined to produce a composite score.

Following the discussion of the likely problems in personality measurement, the present investigation has attempted in this study to examine systematically the implications of particularly the third issue, viz. method variance i.e. variance resulting from method of testing or format of items, as delineated and discussed in next chapters. Before the specific problem of study is presented, the immediately next chapter has been started to review some of the relevant studies already made in this area.
