GENERAL REVIEW.

METHODS OF ESTIMATING INTELLIGENCE AND PERSONALITY AND THEIR APPLICATIONS.

By J. M. BLACKBURN, Ph.D., B.Sc.(Econ.),
Psychologist, Maudsley Hospital.
(From the Central Pathological Laboratory.)

Synopsis.

Part I: Intelligence Tests.

1. THE BINET APPROACH.
2. THE SPEARMAN APPROACH.
3. COMPARISON BETWEEN THE BINET AND THE SPEARMAN APPROACHES.
4. INTELLIGENCE TESTS AS BATTERIES RATHER THAN AS SCALES.
5. NON-VERBAL, PERCEPTUAL AND PERFORMANCE TESTS OF INTELLIGENCE.
6. TESTS FOR VERY YOUNG CHILDREN.
7. DRAWING TESTS OF INTELLIGENCE OR SCHOLASTIC ATTAINMENT.

Part II: Other Methods of Measuring Personality Characteristics.

1. THE QUESTIONNAIRE.
   - (a) *Precautions Taken When Drawing Up Questionnaires.*
   - (b) *Methods of Recording Answers.*
   - (c) *Validity and Reliability.*
   - (d) *Types of Questionnaire.*
2. RATING SCALES.
   - (a) *Different Methods of Giving Ratings.*
   - (b) *Fineness of Discrimination in Ratings.*
   - (c) *Reliability of Ratings.*
3. THE FACTORIAL METHOD OF STUDYING PERSONALITY.
4. THE STUDY OF EXPRESSIVE MOVEMENTS.
5. THE FREE ASSOCIATION METHOD.
6. THE RORSCHACH METHOD.

Part III : Some Applications of Tests.

THE present report attempts to give a summary of some of the ways by which intelligence and other personality characteristics have been tested and measured. The work it describes is scattered through numerous publications and papers, and it is hoped that collecting them together will be advantageous both to those clinical workers who want to apply some of them and to students who need a rapid survey of what has been attempted.

The testing technique is undoubtedly most fully developed in so far as intelligence tests are concerned, and many of the following pages are devoted to describing them. It is hoped, however, that sufficient space has been left in the second section to prevent the description of other measures of personality characteristics from being merely a catalogue. The third section is concerned with applications.

It is an unfortunate necessity that in selecting the tests and methods for description much interesting and important work has had to be omitted. The choice that any one person makes is bound to be different from that of any other, and I can only hope that my selection has been fairly representative. I have, however, added a list of general references at the end, so that anyone who wishes may find other selections and further descriptions of tests.

I should like to express my thanks to Prof. F. L. Golla, at whose instance the report was undertaken, and also to Dr. P. E. Vernon, for their helpful suggestions and criticisms.

I. Intelligence Tests.

1. The Binet Approach.

In 1905 Binet published a scale of questions based on some of the different kinds of practical situation that children might be expected to encounter in the course of their ordinary everyday lives. Binet's original scale was devised for the purpose of diagnosing mental deficiency among the children in Paris schools;* he revised the scale in 1908 and 1911. Terman† extended and standardized the scale for American children (the "Stanford-Binet"), and Burt‡ did the same in this country (the "Burt-Binet"). Burt also proposed a revision of the Stanford-Binet for use in this country (the "Burt-Stanford-Binet"),§ and Terman and Merrill‖ have recently published a complete revision and re-standardization of the Binet scale (the "New Stanford-Binet") which has not yet been standardized in this country.

The method of obtaining a scale was to devise a series of representative questions and then to put them to a large number of children of different ages. The relative ease or difficulty of the questions could be ascertained by finding the percentage of children who could pass each test. The questions could thus be arranged in order of difficulty, and, further, they could be arranged in "mental ages", according to the age of the children who could pass them. Then, when some questions had been eliminated and others added, a scale was finally obtained which comprised several tests for each age. The scale could be applied to other children, and their results compared with the results given by the children of the same chronological age on whom the test had been standardized.

By way of illustration, a few examples are taken from the "Burt-Stanford-Binet". A child of 4 should be able to repeat correctly three digits which are read out to him; a child aged 6 should be able to define correctly in terms of use at least four out of the following six things: chair, horse, fork, doll, picture, table; a child aged 8 should be able to count backwards correctly from 20 to 1; a child aged 11 should be able to say correctly what is absurd about three out of five statements similar to the following: "A soldier writing a letter to his mother started like this, 'Dear Mother, I am writing this with a sword in one hand and a pistol in the other'"; and a child aged 15 should be able to repeat backwards six digits that are read out to him.

In re-standardizing the test on different groups of children, as, e.g., Terman and Burt have done, different age-assignments from the original ones of Binet have sometimes had to be made. This is partly due to the translation of the

* A. Binet, *Année Psychologique*, 1905, xi.
† L. M. Terman, *The Measurement of Intelligence*. Stanford University Press, 1916.
‡ C. Burt, *Mental and Scholastic Tests*. Second edition. London: P. S. King, 1922.
§ Unpublished.
‖ L. M. Terman and M. A. Merrill, *Measuring Intelligence*. London: Harrap, 1937.

French instructions into English idiom, which may alter the difficulty of the test, and partly, perhaps, to slight differences in rate of mental development between the different groups. Thus Burt found that, in comparison with other children tested by the scale, the child of the London elementary school appeared to be somewhat precocious.

Out of Binet's work with intelligence tests arose the very important concept of *mental age* ("M.A."). The average child of nine should theoretically be able to pass all the tests up to and including those for age IX, but should fail in every test for a later age. In practice this is rarely found to be the case, the child sometimes failing on a few tests below, and succeeding on a few above. In order to estimate his mental age, therefore, one credits him with the mental age corresponding to the highest year at which he passes all the tests, adds the appropriate fraction of a year for tests passed in higher ages and subtracts the appropriate fraction for tests failed in lower years. Thus, if a child passes all the tests for age VIII, three out of six for age IX, one out of six for age X, and none for higher ages, he will have a mental age of VIII years plus VI months plus II months = VIII years VIII months.

In order to make the comparison between children of different mental ages easy, Stern and Terman subsequently proposed the adoption of the term "Intelligent Quotient" ("I.Q."). This is obtained from the fraction $\frac{\text{mental age}}{\text{chronological age}} \times 100$. Thus a child of 8 with a mental age of X will have an I.Q. of 125, which is the same as that of a child of 10 with a mental age of XII years VI months, and a child of 10 with a mental age of VIII years VI months will have the same I.Q. (85) as a child of 13 with a mental age of XI.

When a child is re-tested after a lapse of time it has been found that there is usually a fairly close agreement between the I.Qs. obtained in the first and second tests. Thus, a child with an I.Q. of 125 at the age of 8 will be found to have approximately the same I.Q. when he reaches the age of 10 or 12, provided that in both tests he had adequate incentives, and was not prevented by emotional or temperamental factors from doing his best. While his mental age increases as he grows older, his intelligence *level* remains the same. The growth of mental age continues until he reaches the age of between 14 and 16, although more slowly as this age is approached, after which it stops, even if there is a subsequent increase in experience, skill and knowledge. In order to estimate the I.Q. of an adult, therefore, his mental age, as measured by the Binet scale, is divided, not by his actual age, but by the age at which mental growth is supposed to stop. By some testers this is taken to be 16, by others 14.

It should be said, however, that the Binet test is not very satisfactory when used on adults. Apart from the fact that there are far too few tests in the higher mental age-levels, the form in which some of the questions are worded is apt to appear childish to the average adult, who may react in a contemptuous

rather than in a co-operative way. Fortunately the vocabulary test, estimated by Terman to be worth at least three of any of the other tests in the Binet scale, is very appropriate for testing adults. It consists of a series of words of gradually increasing difficulty, which the testee has to explain. The words, that is to say, have to be described in other words, and the relationship of their meaning to that of other words has to be explained. The vocabulary test, or any other intelligence test, does not tell one the *nature* of intelligence, any more than an ammeter tells one the nature of electricity. It does, however, afford a means of indicating the amount of the matter measured that is present. All that is necessary to make the vocabulary test a good test of intelligence is to select the words carefully enough so as not to favour any one special line of experience more than others, and to arrange them in a scale of increasing difficulty.

In the New Stanford-Binet scale some of the shortcomings of the older Binet scales and revisions have been eliminated. Many new tests have been introduced, and the scale has been extended both for the higher and the lower mental ages. In scoring the tests at the upper end of the scale each test passed is given greater weight than are tests passed in other parts of the scale. Account is taken of the decreased rate of growth of intelligence between the ages of 13 and 16 years by decreasing the chronological age ("C.A.") divisor at these ages: for example, a C.A. of 13 years 6 months is taken to represent a divisor of 13 years 4 months, one of 14 years 6 months is taken to represent a divisor of 14 years 0 months, and so on. For all adults a divisor of 15 years 0 months is used.* The result is that the I.Qs. for the upper age-levels are more nearly comparable with the I.Qs. for the lower age-levels than they were in the original scale and its earlier revisions. It is possible for an adult to obtain an I.Q. of 152 on the new scale, compared with a maximum of about 130 on the old scale. An equivalent form of the scale is also provided, either for purposes of re-testing, or for use when coaching in the tests is suspected.

2. The Spearman Approach.

Group tests of intelligence were developed principally in the United States, and the majority of them are not selected according to the statistical principles which govern the selection of tests of the Spearman type.† Spearman was led to the conclusion that "all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or

* See L. M. Terman and M. A. Merrill, *op. cit.*, pp. 29–31.

† The best known of the American group tests are the National, the Otis and the Terman (all obtainable from G. Harrap & Co., London). In this country the best known are probably Thomson's Moray House Tests (obtainable from Prof. Thomson, Dept. of Education, Moray House, Edinburgh), Thomson's Northumberland Mental Tests, Richardson's Simplex Tests, and Cattell's Group Tests (all obtainable from Harrap & Co.), and the group tests of the National Institute of Industrial Psychology (Aldwych House, London, W.C. 2).

specific elements seem in every case to be different from that in all the others ".* He proposed to name this common function " g " and to abandon the term " intelligence ", which had " so many meanings that finally it has none ". Whereas " g " is common to all intellectual processes, " s " (the specific element) is specific to each process. Spearman was led to the formulation of what is known as the " two factor " theory by the observation that when a number of individuals did a series of mental tests, and the correlation co-efficients between their scores on the different tests were worked out, the correlations tended towards a peculiar arrangement known as the " hierarchical order " that could be expressed by a definite mathematical formula known as the " tetrad equation ".

Thus to take Knight's hypothetical example,† suppose we give five tests : Inferences, sentences, numbers, synonyms and codes, and find that the highest correlation coefficient between the scores of the individuals on any two of the tests is, say, ·42, between inferences and sentences, while the correlation coefficient between inferences and the other three tests is ·35, ·28 and ·21 respectively. Then, if we form a table showing the correlation between each test and every other test, of which the first column and the first row represent the above correlations, it will be found that the intercorrelations in every other column and row tend to fall in a descending order :

	Inferences.	Sentences.	Numbers.	Synonyms.	Codes.
Inferences	—	·42	·35	·28	·21
Sentences	·42	—	·30	·24	·18
Numbers	·35	·30	—	·20	·15
Synonyms	·28	·24	·20	—	·12
Codes	·21	·18	·15	·12	—

This hierarchical order occurs when the cross products of *any* square block of four correlation coefficients is equal, within certain defined limits of error, e.g. $(\cdot35 \times \cdot24) = (\cdot30 \times \cdot28)$, or $(\cdot42 \times \cdot12) = (\cdot28 \times \cdot18)$. In actual practice, of course, a *perfect* agreement such as this is not obtained. Expressed algebraically, where a, b, c, d stand for the four tests, and r_{ab}, r_{ac}, etc., stand for the coefficients of correlation between tests a and b, b and c, etc., then $(r_{ab} \times r_{cd}) - (r_{ad} \times r_{bc}) = 0$. This is known as the " tetrad equation ".

Applying mathematical analysis to this phenomenon Spearman concluded that if intellectual ability is due to the presence of two factors, one general and the other specific, then the tetrad equation must be satisfied within certain defined limits of error, and conversely if the tetrad equation is satisfied, then the ability is divisible into two factors, one general and the other specific. It is

* C. E. Spearman, *Amer. Journ. Psychol.*, 1904, xv, p. 202. See also C. E. Spearman, *The Abilities of Man*. London : Macmillan, 1927.

† R. Knight, *Intelligence and Intelligence Tests*. London : Methuen, 1933.

the converse theorem which is applied in two factor analysis, and which has been subjected to criticism by Thomson* and others.

The establishment of the tetrad equation criterion led Spearman and his followers to select, and to concentrate upon, those tests which most nearly satisfied the criterion. In this respect the basis for the selection of tests of the Spearman type differs from the basis used in selecting tests of the Binet type, and also from that used in selecting other group and individual tests of intelligence.

The present tests used as a measure of "g" are almost entirely of the paper-and-pencil type. Spearman's own test† is given orally with written answers, but others, e.g., R. B. Cattell's,‡ are entirely paper-and-pencil. Paper-and-pencil tests have the advantage over tests of the Binet type that they can be given to a group of individuals at the same time, that is to say they may be used as group tests, with a consequent saving of time. The "g" type tests share with other group tests the following characteristics. The subject is presented with a booklet containing a number of different tests, and there are usually one or two examples of what he is required to do at the head of each. Little or no writing is involved, which eliminates the effect on the score of different writing speeds; the subject is merely asked to underline the word representing the correct solution to the problem, out of three or four possible answers. The scoring of the test is thus made objective, and it is not necessary to give partial credits, as it might be if the subject had to invent the correct answer instead of selecting it. All that the tester has to do is to see that the time limits, if any, are observed, although they are often quite generous, and then to correct the answers, a task made easy by the provision of a key.

Some examples of the kind of test used to measure "g" are given below. They are represented in a very simple form, but there is almost no limit to the complexity they may attain:

(a) *Analogies.*—*Black* is to *white* as *big* is to HEAVY, BRIGHT, LITTLE, MAN.

(b) *Synonyms and antonyms.*—*Ill* means the same or nearly the same as BEAUTIFUL, FINE, BRIGHT, UNWELL. *Good* is the opposite of WHITE, BAD, TINY, FALSE.

(c) *Mixed sentences.*—Through air fly birds the: TRUE, FALSE.

(d) *Classification.*—The subject is required to underline the word which is in a different category from the others: CHAIR, TABLE, WHEAT, RUG, CURTAIN.

* W. Brown and G. H. Thomson, *The Essentials of Mental Measurement.* Third edition. Cambridge University Press, 1925. See also G. H. Thomson, *Brit. Journ. Psychol.*, 1935, xxvi, pp. 63–92.

† C. E. Spearman, *A Measure of Intelligence.* First edition. London: Methuen, 1925.

‡ Obtainable from Harrap & Co., London.

(e) *Number series.*—The subject is required to pick out the two numbers which continue the following series: 1, 3, 5, 7, 9, . . . 10, 11, 12, 13, 14.

(f) *Codes.*—A message has to be transcribed according to some principle, such as that each letter must be represented by the letter that precedes it in the alphabet.

(g) *Inferences.*—Jack is taller than Bill, but not so tall as Tommy. Which is the tallest? JACK, BILL, TOMMY.

3. COMPARISON BETWEEN THE BINET AND THE SPEARMAN APPROACHES.

It is useful to compare the "g" test situation with that of the Binet test. The "g" tests are based on mathematical reasoning following on the observation that the tetrad equation is satisfied in a number of tests, and they are selected with a view to satisfying statistical rather than psychological criteria as nearly as possible. The Binet tests, on the other hand, are based on assumptions concerning the kinds of practical problems that individuals are likely to encounter in the course of their ordinary lives, and they are selected with a view to approximating closely to such situations. As a result of this difference in the basis of the selection, it is difficult to know how far the two types of test are measuring similar things, until there is an agreed interpretation of "g" in psychological terms. Yet Spearman's own proposition that the term "intelligence" has been given so many meanings that finally it has none would seem to offer an insuperable difficulty to the interpretation of "g". Different interpretations might similarly endow "g" with so many meanings that finally it had none. It cannot, however, be maintained that the Spearman school has lived up to its principles in this respect. Spearman names his own test a "measure of intelligence", and in his books he continually refers to intelligence. The unfortunate result is that the casual reader is apt to receive the mistaken impression that the interpretation of "g" in psychological terms is supported by as much statistical and mathematical evidence as the general factor theory. It is, however, important to point out that though "g" may mean something mathematically, it means nothing psychologically or clinically until it has been interpreted.

The "g" tests, in common with other group tests of intelligence, differ from the Binet tests in method of presentation. In the "g" type the testee is given a booklet in which he is instructed to draw lines under appropriate words. Even when the test is given as an individual test the testee is encouraged to do his best in exactly the same way as every other testee; there is no modification to suit temperamental differences. This procedure is approved of on the grounds of the desirability of standardization. It is also held that when the tests are given as group tests the "social stimulation" of the group acts in many cases as an incentive.

In giving a Binet test, on the other hand, the testees can be encouraged individually in the manner considered most likely to induce them to exert themselves to their maximum capacity. This method is thought to have an advantage in that the final result provides a better indication of the individual's true capacities than if all had been presented with the same objective situation. It is claimed that in the Binet situation it is often possible to estimate immediately whether a low score is due to inhibiting emotional factors or to an innate lack of intelligence. In the "g" test situation, and in other group tests of intelligence, this can be estimated only after further investigation.

The Binet situation, however, makes more demands on the examiner. The high degree of objective standardization of the "g" and other group tests is intended to make the giving of the tests foolproof, so that reliable scores may be obtained, even when the tests have been given by examiners with minimal testing experience. If the Binet tests are to be reliable, on the other hand, they must be given by examiners who are not only experienced, but also capable of adopting an understanding and sympathetic attitude towards the different kinds of children they have to test. The kind of encouragement that is successful with one child may fail completely with another; the tests which are most interesting to one may prove boring to another. With younger children the testing can often be approached through toys, the tests being given in the form of new games. With older children Burt and others recommend that the initial test should be a shock absorber, selected because it unfailingly serves to interest the child.*

In preparing children for school examinations coaching is often employed. Coaching, however, defeats the purpose of an intelligence test examination. The object of both group and individual tests is to provide different children with conditions in which their true abilities can be ascertained, and the effect of coaching is simply temporarily to nullify this purpose. Practice tests, based on similar principles, but not identical with, the tests that are to follow are often provided in the group test situation to act as shock absorbers, much as in the Binet situation the initial test is often selected for the same purpose. Practice tests are, however, quite different in effect from specific coaching in the actual questions that are subsequently to be presented. The Binet situation has the advantage over the group test situation that coaching can more easily be discovered. The greater the clinical experience of the examiner the easier is it for him to differentiate between an answer that is genuine and one that is the result of specific coaching. For instance, suspicions may be aroused by answers that are given instantaneously and monotonously, and in phraseology that is inconsistent with the child's normal mode of expression.

* Many suggestions concerning the administration of the Binet scale, and the precautions that it is necessary to take, will be found in, for example, C. Burt, *Mental and Scholastic Tests*, pp. 9 *et seq.*; L. M. Terman and M. A. Merrill, *Measuring Intelligence*, chap. iv; and E. P. A. Hunt and P. Smith, *A Guide to Intelligence and Other Psychological Testing*. London: Evans, 1937.

Such things do not appear in the group test situation, where a number of children are tested at the same time, and they are also less likely to appear when a paper-and-pencil test is given individually. The point is important also in emphasizing the necessity for considerable clinical experience before the most accurate results may be obtained from a Binet test.

4. INTELLIGENCE TESTS AS BATTERIES RATHER THAN AS SCALES.

The tests so far considered have been designed for presentation as scales. The whole scale, or a large part of it, must be presented before the determination of a person's intelligence is made. However, Kent,* who appreciates the value of the individual question-and-answer method of testing, is of the opinion that none of the editions of the Binet scale, and none of the existing group tests, have yet produced an instrument which is sufficiently flexible to fit in with the interests and tastes of all the different patients that are likely to be encountered in a clinic. Her point is that a composite scale, which has to be used as a whole, may give results which are very unfair to some patients. Furthermore, since the time available for testing in clinical situations is usually very limited, the application of a whole scale may be wasteful, for it involves the application of non-discriminative material, both at the upper and at the lower ends of the patient's range.

She therefore recommends the development of a battery of language tests on the lines of the Pinter and Paterson performance battery.† Items which can be graded in difficulty should be selected, developed into a graded series, and standardized as an independent unit. Each unit should be so graded as to cover the entire range of mental levels for which it is appropriate, but for economy of presentation the standardization should be for overlapping sections rather than for—or in addition to—standardization for the series as a whole.

Then, when a sufficient number of such test units had been developed, an examiner could make up a battery by selecting such units as were best suited to a particular patient. The examination would then be arranged to fit the patient, instead of the patient being held responsible for fitting—or for not fitting—the test.

Such a battery of units would yield a series of independent ratings, and Kent suggests that if it is held necessary to represent a patient's score by a single figure the medium of the ratings should be placed on record for reference.

An additional advantage of such a loose collection of independently standardized units would be that out-of-date items could be dropped, and new units introduced as the occasion required, without disturbing the other units

* G. H. Kent, "Suggestions for the Next Revision of the Binet Scale", *The Psychological Record*, 1937, i, pp. 407–434.

† See below, Section 5.

of the battery; whereas a fixed scale, standardized as a whole, becomes more and more unsatisfactory as some of its items pass out of date.

Kent's proposal is of considerable importance, and it is to be hoped that it will at least be given a trial in this country.* Any worker may contribute to the development of a battery by helping to standardize some test item or items in which he is particularly interested. The final result would, of course, be an instrument that diverges even more sharply than the Binet scale from the objective similarity of group tests or Spearman type tests. The proposal, however, is defended on the grounds set out above, and on the further assumption that intelligence cannot be measured as an entity, but only as a type of behaviour. No physician who is asked to give a report on a patient would think of expressing his result in terms of a "physical quotient", and no psychologist, according to this view, should be content to express his report in terms of an intelligence quotient. Mental characteristics are at least as variable as physical. The comparison between individuals should be based on differences in their behaviour, judged by the type of test in which they excel or fail, not on their possession of a few points more or less of an abstract entity.

5. Non-verbal, Perceptual and Performance Tests of Intelligence.

In order to eliminate the language factor which enters into the tests so far described, various attempts have been made to produce intelligence tests of a non-verbal type. In the American Army tests used during the war two scales were provided, one of which (the "Army Beta") was in non-verbal form and was used for testing men of foreign descent, as well as illiterates.† In this country one example is Sleight's non-verbal intelligence test,‡ for testing children between the ages of 6 and 10 years. It is designed as a group test, and consists of ten tests, of the substitution, classification, series continuation and other types, arranged in the form of pictures or drawings. For instance, instead of the words "TRUMPET, TABLE, VIOLIN, DRUM", pictures of these objects are given and the child is told to mark a ring round the object that is in a different class from the others. The principles on which the test are based are the same as have been previously discussed, but the form in which they are presented is different. R. B. Cattell also includes a few non-verbal items in his mainly verbal group tests of intelligence, scales 2 and 3.

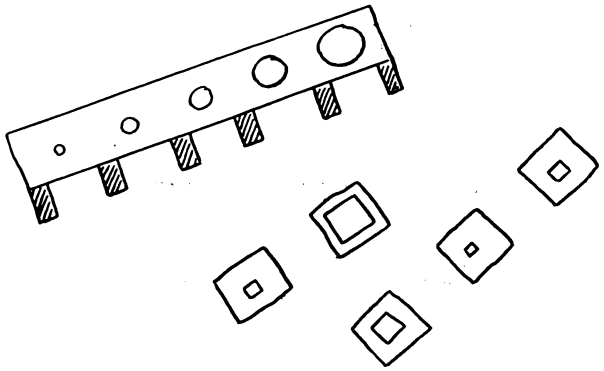
A scale which is based on the matching technique, and is as far as possible independent of cultural background, of the time element, and of pantomimic

* Kent has published her own preliminary battery of this type—see G. H. Kent, *Journ. Genetic Psychol.*, 1934, xliv, pp. 49–68—but it has not yet been adapted for use in this country, or standardized over here.

† See C. S. Yoakum and R. M. Yerkes, *Mental Tests in the American Army*. London: Sidgwick & Jackson, 1920. Another American non-verbal test is the Otis Primary, which has been modified for use in this country by Lewis and Burt, and is obtainable from Harrap & Co.

‡ Obtainable from Harrap & Co.

instruction to the testee, has been devised by Leiter.* It has been carefully standardized and used for comparing the intelligence of different races, e.g., the Chinese and Japanese in Hawaii. It consists of two tests for each year between the ages of V and X, and one for each year from XI to XVI. A shorter form of the scale was used by Porteus when testing African natives. The principle of the test is the matching of appropriate designs. In a test for year VII, for example, a strip on which five graduated circles are drawn is placed before the subject, and he is also given five blocks, on each of which is drawn a square of different size. The subject merely has to arrange the squares in a graduated series, corresponding to that of the circles drawn on the strip. This test, and another from the same scale, are illustrated below :



Test for age VII, 1. *Size gradations.*

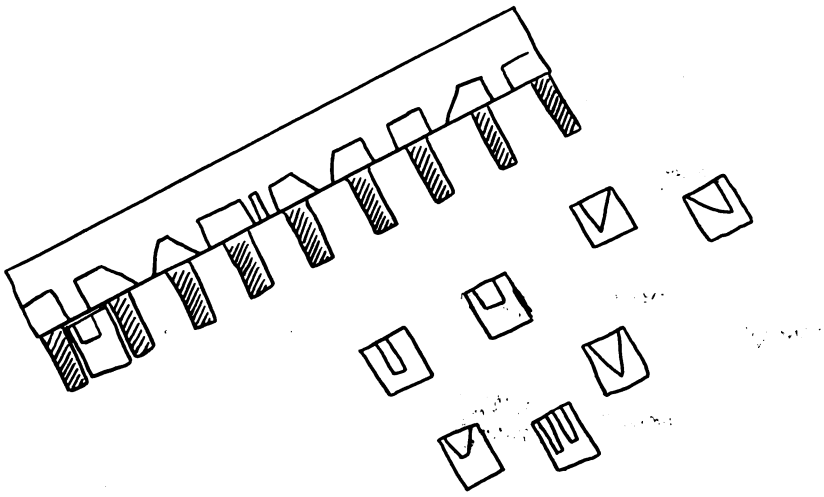
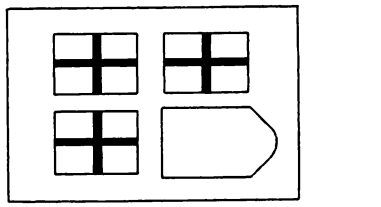


FIG. 1.—Test for age XI. *Line completion.*

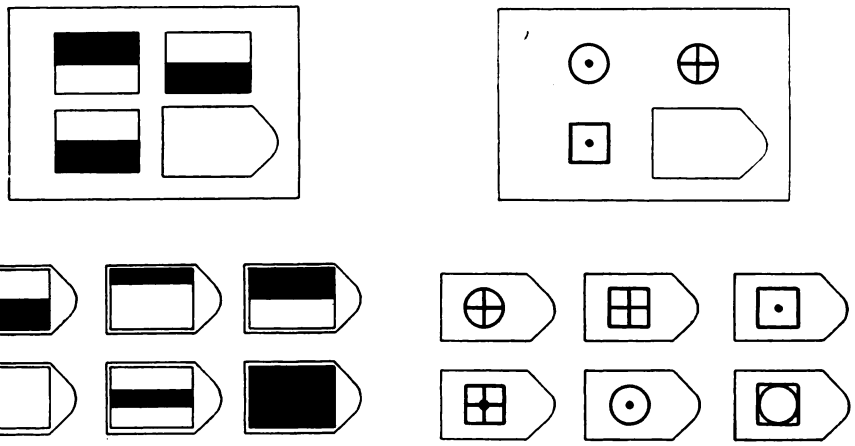
Examples of tests from the Leiter International Performance Scale.

* R. G. Leiter, "The Leiter International Performance Scale", *University of Hawaii Research Publications*, 1936, XV.

Different principles are involved in the perceptual tests devised by Penrose and Raven.* This series is based on the Analogies type of verbal test, and the general structure may contain four elements, nine elements, or be even more complex. The relationship may be one of (1) similarity, (2) opposition or (3) addition, as illustrated below. The testee is presented with a board on which there are three figures and a blank space; he is also given a number of loose blocks on which are printed different designs. He has to select the block



Example 2.



Example 3. Example 4.

FIG. 2.

* L. S. Penrose and J. C. Raven, *Brit. Journ. Psychol.* (Medical Section), 1936, xvi.

https://doi.org/10.1192/bjp.84.353.1008 Published online by Cambridge University Press

bearing the correct design and insert it in the lower right-hand corner of the board, so that the analogy between the designs is completed.

A number of other tests of a non-verbal nature go by the name of "performance tests of intelligence". They were originally devised to test deaf children, illiterates and foreigners. They are, however, of general value in that they test an individual's intelligence in concrete situations. Although the results obtained from performance tests often show close agreement with those obtained from verbal tests, there is sometimes a considerable difference between them, an individual who shines in one type making a poor score in the other. Even in the Binet scale there are not enough concrete situations, and it is wise to supplement the scale by making use of one or two performance tests.

Full descriptions of performance tests may be found in the books of Pintner and Paterson,* Arthur,† or Drever and Collins.‡ All these authors provide complete scales for the measurement of intelligence, based on a large number of different performance tests. Pintner and Paterson, and Arthur, also present norms for each test item standardized separately, with the consequent advantage that their scales may be alternatively used as batteries containing a larger or smaller number of items.§ A shorter scale is that of Alexander.‖ It consists of three tests only, the passalong, which Alexander devised himself, the cube construction and the Kohs' blocks design test. The last two are presented in the modified form adopted by Drever and Collins, although the method of scoring is different. It is probably the most satisfactory scale for adults and for children over the age of ten, but the tests are too difficult for younger children.

Even when there is not enough time to give a whole scale, the inclusion of one or two separate performance tests, particularly if they are sufficiently difficult, is often invaluable, not only because they tend to balance the results of tests with a verbal bias, but also because of the indications they provide of temperamental characteristics. From the way the individual does the tests one may discover, even better than in the Binet test situation, how he reacts when working under pressure, whether he is inhibited or aggressive, quick or slow, co-operative or negativistic, self-conscious or assured, and easily affected by praise, either advantageously or deleteriously, or unaffected by it, and so on. Among the best known and most frequently applied performance tests are Kohs' blocks design, Healy picture completion No. 1 and No. 2, the Seguin-Goddard form-board, cube construction, cube imitation and the Porteus mazes. These are all described in detail in the works already referred to.

* R. Pintner and R. Paterson, *A Scale of Performance Tests*. New York: Appleton, 1917.
† G. Arthur, *A Point Scale of Performance Tests*. New York: Commonwealth Fund, 1930.
‡ J. Drever and M. Collins, *Performance Tests of Intelligence*. Second edition. Edinburgh: Oliver & Boyd, 1936.
§ See above, Section 4.
‖ W. P. Alexander, "Intelligence, Concrete and Abstract", *Brit. Journ. Psychol. Monog. Suppl.*, 1935, No. 19.

Performance tests have, however, certain disadvantages. Many are expensive, they take a long time to do, and norms which are available in this country are still for the most part inadequately standardized.*

6. Tests for Very Young Children.

The Binet scale and its earlier revisions are not considered very satisfactory for testing the intelligence of children under the age of about 5 years. Other scales are often employed. The position is altered now, however, for the new Stanford-Binet scale includes tests for half-year intervals between the ages of II and V as carefully standardized as the tests in other parts of the scale, and made more interesting and attractive to younger children than were the tests in the lower age-grades of the former Binet scales.

R. B. Cattell devised some tests, based on Spearman's principles, for children of ages ranging from 4 to 8 years,† but it is often necessary to test the intelligence of children younger than this, and many clinics use for this purpose the Gesell developmental schedule or the Merrill-Palmer scale. Gesell‡ presents over forty tests arranged in motor, language, adaptive and personal-social behaviour groups, providing tests for the "neonate", 4, 6, 12, 18, 24, 36, 48 and 60 months' old child. Exact scores are not given, but the performances are classified according to the frequency with which each item of behaviour has been found to occur at a definite age-level. When the various items are scored on this basis it is possible to obtain a general idea of a child's stage of development. Gesell has intentionally arranged the scoring of the tests so as to avoid giving an exact I.Q.

The Merrill-Palmer scale§ consists of 38 tests selected from among 79 and providing 93 separate diagnostic items. The range of the scale is from 18 to 72 months. The tests are of different types, including language tests, all or none tests, form-boards and picture tests, and other tests of motor co-ordination. The different types of test are scattered through the scale, which is made to approximate closely to a play situation, although actual performance is standardized as soon as full co-operation has been obtained from the child.

* The norms employed most frequently in this country may be found, for a number of performance tests, in F. Gaw, "Performance Tests of Intelligence", *Industrial Fatigue Research Board Report No.* 31, London: H.M. Stationery Office, 1925. Some clinical workers are not, however, entirely satisfied with these norms, and prefer to use the American norms given by Pintner and Paterson or by Arthur (*op. cit.*). New norms for some of the tests, based largely on child guidance material, may be found in P. E. Vernon, *Brit. Journ. Educ. Psychol.*, 1937, and R. B. Cattell, *A Guide to Mental Testing*, London: University of London Press, 1936, gives fresh norms for some of the tests mentioned. There is, however, a great need in this country for a complete re-standardization of most of the tests over a wide age-range.

† R. B. Cattell, *A Guide to Mental Testing.* London: University of London Press, 1936.

‡ A. Gesell, *The Mental Growth of the Pre-School Child.* New York: Macmillan, 1925.

§ R. Stutsman, *Mental Measurement of Pre-School Children.* New York: World Book Co., 1931. English norms by Hilda Bristol have been published in Prof. Hamley's section on Mental Tests in the *Education Year-Book* (1935). These norms are about five points stricter than the American norms over the whole range. Some clinicians still prefer to use the American norms.

Tests are marked "passed" or "failed", "omitted" or "refused", the last category designating those in which the tester failed to gain the child's co-operation. In estimating the final score, refused tests do not necessarily count against the child, as they do in some scales which involve a rigidly standardized procedure.

7. Drawing Tests of Intelligence or Scholastic Attainment.

Goodenough* devised a technique for measuring intelligence by asking children to draw a picture of a man. The resulting drawing is subjected to a detailed analysis, and scored on the basis of what is included in or omitted from the drawing. The final score gives an indication of the child's mental age, and the drawings may be interpreted qualitatively as well.

Other writers prefer to regard drawings as indications of scholastic attainment rather than of intelligence, the child's drawing being compared with median samples for different age-levels, and used, together with other scholastic tests, e.g., of reading, arithmetic, spelling, composition, and so on, in diagnosing educational retardation.†

II. Other Methods of Measuring Personality Characteristics.

1. The Questionnaire.

One of the most popular methods of investigating personality differences is a questionnaire. It is widely used, particularly in the United States, for little labour appears to be rewarded by quick returns. The accuracy of questionnaire results, however, is much more affected by the goodwill of the person who fills it up than is the accuracy of a test result. In a test it is more difficult to affect one's score by giving false reactions than it is in a questionnaire, for tests are intended to find out what a person *can* do, whilst questionnaires are in most cases intended to find out what he had done, or how he thinks or feels about different problems.

(a) *Precautions Taken When Drawing Up Questionnaires.*

It sometimes happens that a person is able to give the correct answer to a questionnaire, but does not do so. For instance, Persing‡ made a comparison between what students said they would do if asked to report errors in the marking of their papers and what they actually did. He made systematic errors in grading their examination papers, marking some too high and some too

* F. L. Goodenough, *The Measurement of Intelligence by Drawings*. New York: World Book Co., 1926.

† C. Burt, *Mental and Scholastic Tests*. Second edition. London: P. S. King, 1922. See also R. B. Cattell, *A Guide to Mental Testing*. London: University of London Press, 1937.

‡ K. M. Persing, "Morals and Chemistry", *Educational Review*, 1926, lxxii, pp. 164–168.

low, after having asked the students to report any errors they might notice. 97% of the low marks were reported, but only 9·5% of the excess marks. Later on the same students were asked to say what they would do if they found that they had been given marks that were too low or too high, and 80·5% said they would report marks that were too high—an interesting comparison with the figure actually found.

In some forms of questionnaire, however, answers which are given purely for effect are as useful as truthful answers, for instance, in questionnaires which indicate a person's adjustment or maladjustment.*

The frankness with which a questionnaire is answered also depends to a large extent on whether it deals with matters which are affected by moral, social or religious prejudice, and if necessary a check may be kept on truthful or untruthful answers by introducing some of the following techniques:

(1) If the questionnaire is long enough the same question may be asked in slightly different forms in different places, the subject's replies being subsequently compared.

(2) Sometimes one can infer general tendencies apart from the truth of specific answers.

(3) The answer and the interpretation of the answer may be different. For instance, in Hartshorne and May's "Character Education Inquiry"† there was a series of thirty-six questions, such as—

> 8. "Have you ever disobeyed any law of your country or rule of your school?"
>
> 25. "Did you ever pretend that you did not hear when someone was calling you?"
>
> 29. "Do you read the Bible every day?"
>
> 32. "Do you usually correct other children when you hear them using bad language?"

The questions were given to a number of school-children and it was found that the children who answered "No" to the first two and "Yes" to the last two, and answered the thirty-two other questions in a similar way, were likely to cheat *more* in classroom situations, with money and at games, than were the children who answered them in the reverse manner.

(4) Questionnaires may be devised so as to test something different from what they appear on the surface to be testing, e.g. Terman and Miles's questionnaire to indicate degree of masculinity or feminity is called an "attitude-interest analysis test".‡ Such a technique may be particularly appropriate in this country, where people's co-operation in filling-up the questionnaire is more difficult to obtain than it is in America.

(5) A rather similar technique is used by Watson in his "Survey of Public

* See below, p. 1027.

† H. Hartshorne and M. A. May, *Studies in Deceit*. New York: The Macmillan Co., 1928.

‡ L. M. Terman and C. C. Miles, *Sex and Personality*. New York: McGraw Hill, 1936, pp. 600.

Opinion ".* It is claimed that a person may betray his prejudices more easily when he is asked to give an opinion about a situation that is immediate and personal than when he is given the same situation in a remote and impersonal form. Similarly, his bias on religious and economic issues may lead him to neglect the logical strength of arguments propounding views that are different from his own, and to over-estimate the strength of arguments which expound views with which he is in agreement. Watson's questionnaire has, therefore, been developed so as to apply these principles unknown to the subject.

(6) Precautions must be taken with the wording of the questions, and suggestiveness must be avoided. Muscio† has shown that some types of question, for instance, those phrased in a form like "Did you see the . . . ", or "Didn't you see the . . . " are highly suggestive, whereas others, like "Was there a . . . " are not. In general Muscio found that the most reliable form of question is one given a subjective direction, but which contains neither a negative nor the definite article. Again, if the questionnaire deals with matters having a social, moral or religious connotation, the questions which suggest a conventionally acceptable alternative are likely to be more frankly answered than questions which suggest the opposite. For instance, Vernon suggests that more people may be willing to answer "No" to the question "Have you always declared the whole of your income to the Inland Revenue Commissioners?" than will answer "Yes" to the question, "Have you ever defrauded the Inland Revenue Commissioners by omitting to declare part of your income?"

(b) *Methods of Recording Answers.*

One method of recording answers is the "true-false" or the "yes-no" technique. Statements are written down and the respondents have to underline the appropriate answer. For instance:

"I read the Bible every day"—YES ; NO ; or,

"I read the Bible every day"—TRUE ; FALSE.

Sometimes an intermediate category is introduced in case the respondent is doubtful.

A second method is the *multiple choice* method, in which a number of alternative answers are offered and the respondent chooses one by underlining it, e.g.:

"I read the Bible
- Every day,
- Twice a day,
- Once a week,
- Once a year,
- Never."

* G. B. Watson, "The Measurement of Fair-Mindedness", *Teachers' Coll. Contrib. to Education*, 1925, No. 176, pp. 97.

† B. Muscio, "The Influence of the Form of a Question", *Brit. Journ. Psychol.*, 1918, viii, p. 351.

Rating scale techniques have also been employed, but they will be referred to in Section 2 of this chapter.

If possible, space should always be left so that the respondent can add any comments or qualifications that may help to clarify his point of view.

(c) *Validity and Reliability.*

A questionnaire is valid if it measures what it sets out to measure; it is reliable if respondents tend to give the same replies when the questionnaire is repeated. It is possible to have a questionnaire with high validity but low reliability, for example, a questionnaire might, at a given moment accurately measure what it sets out to measure, but, on account of rapid fluctuations in the respondent's opinions, not give the same result when repeated; or, again, the questions in the questionnaire might be truly representative of the qualities it was desired to measure, but because the questionnaire was too short, or because the respondent intentionally desired to deceive, the picture that emerged from the answers might not be sufficiently detailed or accurate. It is also possible to have a questionnaire with high reliability but low validity—for example, one which gives identical results on repetition, but results which are not a true indication of the characteristics it is intended to measure.

One method of making a questionnaire valid is to select questions on *a priori* principles, choosing only those that are consistent with a preliminary analysis of the characteristics to be measured. Another method is to see whether the answers given by extreme groups, for instance, die-hards and revolutionaries, clergymen and atheists, etc., are significantly different.

Assuming that other difficulties have been disposed of, the question of the representativeness of the answers is still important. If it is required to sample the views of a large and heterogeneous population, it is necessary to take many precautions before the questionnaire is issued. Some groups, particularly educated people, give franker and easier co-operation than other groups, but even educated people may be more willing or less unwilling to co-operate if the questionnaire is supported by some well-established institution. Apart from this, however, it is essential to be familiar with the statistics of sampling, so that a truly representative section of the heterogeneous population may be studied. In other cases, where it is required to investigate the views of a more homogeneous group, e.g., college students, the desired results may be more easily achieved.

The reliability of a questionnaire may be studied either by what is known as the " split-half " technique, that is to say, by finding the correlation between the answers to two halves of the questionnaire, or by giving the same or a similar questionnaire at a later date.

(d) *Types of Questionnaire.*

So many questionnaires have been proposed for measuring so many different characteristics that it would be foolish to attempt to describe many of them here.* Those we are about to discuss are selected for purposes of illustration:

(1) *Woodworth's psycho-neurotic inventory and its revisions.*—This consisted originally of 225 questions based on the different types of neurosis described in J. T. MacCurdy's book *War Neuroses*. Some of the questions are intended to investigate attitudes rather than facts, for instance, an affirmative answer to the question, "Do you think tobacco has harmed you?" betrays an anxiety irrespective of the fact whether the tobacco has actually harmed or not. The inventory has undergone many revisions,† one of which is called by Thurstone the "personality schedule".‡ This contains forty-two questions, of which the following are examples:

"6. Are your feelings easily hurt?
"10. Do you worry over possible misfortunes?
"15. Do you get discouraged easily?
"34. Do you lack self-confidence?"

(2) *Pressey's* ×*-o tests.*—A different type of questionnaire is Pressey's cross-out (×-o) test.§ It was put forward for purposes of research, but it has been widely applied. In one section the respondent has to cross out anything he thinks is wrong, e.g., "begging, smoking, flirting, spitting, giggling"; in another he has to cross out anything which makes him feel anxious or worried, e.g., "loneliness, work, forgetfulness, school, blues"; and in a third he has to cross out anything in which he is interested, e.g., "kissing, flirting, pretty girls, talkative girls, athletic girls". A total of 375 items are included.

(3) *Allport's a-s test.*—In the Allport ascendance-submission (a-s) test‖ a number of situations are presented and the respondent has to say what he would do if he found himself in them, e.g.:

> "At church, a lecture, or an entertainment, if you arrive after the programme has commenced and find that there are people standing, but also that there are front seats available which might be secured without 'piggishness' or discourtesy, but with considerable conspicuousness, do you take the seats—
> "Habitually .....
> "Occasionally.....
> "Never....."

* For full descriptions see, e.g., P. M. Symonds, *Diagnosing Personality and Conduct* (New York: Appleton-Century, 1931), or P. E. Vernon, *Industrial Health Research Board Report, No.* 83 (London: H. M. Stationery Office, 1938).

† A revision to be used on children in this country is suggested by C. Burt, *The Subnormal Mind*, Oxford University Press, 1937.

‡ L. L. Thurstone and T. G. Thurstone, "A Neurotic Inventory", *Journ. Soc. Psychol.* 1930, i, pp. 3–30.

§ S. L. Pressey, "A Group Scale for Investigating the Emotions", *Journ. Abn. and Soc. Psychol.*, 1921, xvi, pp. 55–64.

‖ G. W. Allport, "A Test for Ascendance-Submission", *ibid.*, 1928, xxiii, pp. 118–136.

The questionnaire is constructed in two forms, one for men and one for women, but it has not yet been standardized in this country. Exploratory work, however, appears to indicate that the American norms indicate a greater degree of ascendance than is found in comparable groups in this country.

(4) *Thurstone's attitude scales.*—Thurstone and his co-workers have developed about forty scales for measuring different attitudes.* That, for instance, for measuring attitude towards the movies contains forty questions, each given a value between 0 and 4·7, representing very unfavourable and very favourable attitudes respectively. The average, representing indifference, is about 2·5. The following statements represent attitudes that are very favourable, very unfavourable, and indifferent :

"27. (4·7) The movies are the most powerful influence for good in American life.

"15. (0·0) It is a sin to go to the movies.

"14. (2·4) Sometimes I feel that the movies are desirable and sometimes I doubt it."

The figures in brackets represent the degree of favourableness or of unfavourableness of the statement.

The respondent marks the statements with which he is in agreement, thus indicating his average attitude and also the range of opinion with which he is willing to ally himself.

(5) Watson's "Measurement of Fair-mindedness"† is a questionnaire for measuring attitudes, prejudice and impartiality. It employs the novel technique described above in section (*a*) 5.

2. Rating Scales.

Provided that the proper precautions are taken in drawing up and in interpreting questionnaires, useful information about the proportions of people possessing certain views or attitudes, or expressing certain types of behaviour, may be drawn from them.

In some cases, for instance in the Thurstone attitude scales, it is possible to express the results in exact numerical terms. But many questionnaires possess the disadvantage over other methods of studying personality differences that unless steps are taken to hide the true purpose of the questionnaire, the respondents will tend to over-estimate themselves, particularly on the qualities which are considered desirable in their own social culture, and to under-estimate themselves on those which are considered undesirable. This "halo" effect, which must also be guarded against when other methods of assessing personality characteristics are employed, is perhaps most difficult to avoid when

* L. L. Thurstone, "The Measurement of Social Attitudes", *ibid.*, 1931, xxvi, pp. 249–269.

† G. B. Watson, "The Measurement of Fair-Mindedness", *Teachers' Coll. Contrib. to Education*, 1925, No. 176, pp. 97.

the respondent is attempting to describe his own characteristics. For many purposes some form of rating scale is often more suitable than a questionnaire technique.

(a) *Different Methods of Giving Ratings.*

One of the simplest methods is the *ranking* method. A teacher can more easily rank her pupils in order of intelligence from the lowest to the highest, than make an estimate of the amount of intelligence that each possesses.

Another method is to make *paired comparisons*. Every individual is compared with every other individual in the group. Thus *A* may be rated as more intelligent than *B*, less intelligent than *C* or *D*, etc., *B* as less intelligent than *A*, *C* or *D*, etc., and so on through all the possible comparisons. An individual's final score is the number of times he is preferred.

Another technique is the *man to man* method. It was widely used in the American army during the war. Each rater selects five men, one possessing a minimal, another a maximal, and the other three intermediate degrees of the quality being rated. These men he uses as a scale by which to measure all the other men he rates. There is, however, no method of comparing one rater's scale with that of another; what one person thinks is average may be thought by another to be very much or very little.

The same difficulty is encountered in *numerical ratings*. Raters are given scales which range from 0 to 5, or from 0 to 7, or from -3 to $+3$, etc., and they have to rate the amount of the quality that each person to be rated possesses by representing it as a point on the scale.

In order to avoid the difficulty arising from the incomparability of different raters' standards a *graphic rating scale* is more often used. Different degrees of the quality are described in verbal terms by the use of descriptive adjectives. The rater then marks a position on the scale that he thinks the person to be rated reaches. The ratings may be subsequently converted into quantitative scores.

For example, a person might be rated for "personal care" somewhere along the following five-point scale.

Fastidious in dress.	Good taste, neat, clean.	Passable and inconspicuous.	Careless in dress and cleanliness.	Slovenly and unkempt.

In the *guess who* method the rater is given a series of character sketches representing people with different degrees of the quality being rated, and he has to fit an appropriate sketch to each individual. This method also materially reduces the degree of divergence between the standards of different raters.

(b) *Fineness of Discrimination in Ratings.*

It has been claimed that the qualitative methods, e.g., *ranking* and *paired comparisons* possess the advantage over the quantitative methods that the degree of discrimination may be as fine as is desired. However, if the number of people ranked is large, a degree of artificiality is introduced into the rankings. For although the extreme individuals at both ends of the scale may stand out, there will be a crowd of average people in the middle who are not sufficiently different from one another to be accurately ranked in order. The rater will be obliged to make false comparisons between them if he is forced to rank them. In rating scales, too, the question of fineness of discrimination is important. If the scale is too coarse the ratings will be unreliable, because there will not be enough grades into which everyone can be fitted. On the other hand, if the scale is too fine the ratings will be unreliable, because the rater may become fatigued while looking for minor degrees of difference between the individuals he is rating. Symonds* suggests that on the whole seven grades is the most satisfactory number to use, although in practice five are more frequently employed.

(c) *Reliability of Ratings.*

When investigators have studied the reliability of ratings, that is to say the correlation between the ratings of different raters† they have sometimes found that the degree of agreement is rather low. It has, therefore, been suggested that ratings should not be used unless they are based on the average results of at least three independent raters. A single individual's ratings are sometimes affected by what is known as the " halo " effect. This occurs when a person allows his general reaction to the personality of the man he is rating to bias his ratings, with the result that he rates most of his characteristics either too high or too low, according to whether his general impression of the personality is favourable or unfavourable.

3. The Factorial Method of Studying Personality.

When a series of tests are given to a number of individuals and the correlation coefficients between their scores on the different tests do not satisfy the tetrad equation criterion within the errors of sampling,‡ it is held that the explanation is to be found in the fact that factors other than " g " and " s " are involved in the performance of the tests. A number of different factors

* P. M. Symonds, *op. cit.*, p. 79.

† It is well to notice that reliability is measured in a different way in rating scales from the way in which it is measured in questionnaires. A rater may be *consistent* in his ratings, i.e., give approximately the same ratings on different occasions, without his ratings agreeing with those of the majority of other raters.

‡ See Part I, pp. 1010–1023.

have been advanced on these grounds.* Let us take, by way of example, the tests called " p " tests They are of different forms, one type being the following :

(1) The subject writes *abcd*, *abcd*, etc., for 15 seconds, then *ABCD*, *ABCD*, etc., for 15 seconds, then repeats the first activity and then the second.

(2) He writes *aAbBcCdD*, *aAbBcCdD*, etc., for four periods of 15 seconds each.

The " p " score is the number of letters in (1) divided by the number of letters in (2), omitting errors.

It is, of course, very likely that a test such as this measures something rather different from what is measured by the kind of test described in the last chapter which is used to measure " g ". The trouble begins when an interpretation is put on what the tests are measuring, for, as with the " g " tests, the naïve reader may be inclined to believe that the interpretation also rests on a mathematical basis. In this respect the factorial psychologists are not entirely innocent of blame ; and instead of giving their tests non-tendentious names like *alpha*, *beta*, *gamma*, etc., they give them, and use them as if they stood for, the first letters of words which express personality traits—though Spearman long ago gave the warning that terms in common use may bear so many meanings that finally they have none. Thus, as soon as " p " is understood to refer to " perseveration " we have passed (although the factorial psychologists may not tell us as much) from mathematics to qualitative interpretation. It is still more surprising to learn what a varied array of qualities are understood to be included in the name " perseveration ". Thus, a person who makes a low " p " score is, among other things, inclined to be nagging, restless and fussy, inconsiderate and tough, selfish and impetuous, interested in mechanical, scientific and mathematical matters, and to have few dreams ; whereas a person who makes a high " p " score is sceptical and pessimistic, conservative in his habits, sensitive, interested in history, languages and humanities, absent-minded, dreamy and sentimental, and slovenly in dress.† " p " means all these things and is disclosed by such activities as manipulating letters of the alphabet. Confusion rather than clarity would seem to have been added to the picture by calling " p " perseveration. One can imagine that Humpty-Dumpty would have paid a high price for this word which has so many different and unexpected meanings.

This technique of attempting to split personality into a number of independent traits or factors is also open to criticism from a different point of view. Vernon‡ points out that because two individuals are given the same score in a rating scale or a test, this does not necessarily mean that they really possess the same trait. The organization of the personalities of the two individuals

* See R. B. Cattell, *A Guide to Mental Testing*. London: University of London Press, 1936.
† See R. B. Cattell, *op. cit.*, p. 209.
‡ P. E. Vernon, " Can the ' Total Personality ' be Studied Objectively ? ", *Character and Personality*, 1935, iv, pp. 1–10.

may be entirely different, so that a certain amount of a characteristic in one personality organization may mean something different from the *same* amount of that characteristic in another individual with a different type of personality organization. " Attempts at measurement, whether with tests or with ratings, inevitably disrupt the personality into such separate bits as can be handled by our quantitative techniques, but they cannot put it together again."*

Vernon himself used personality organizations as a basis in his investigations of personality, and he has developed it in the matching technique.† He reports an experiment‡ in which photographs were (*a*) matched with brief character sketches, and (*b*) ranked on separate traits, such as intelligence, sociability, emotionality, artisticness, leadership and efficiency. In all cases the photographs were matched with the sketches more accurately than they were ranked on the separate traits.

A point of technique in the use of the matching method is worth considering at this point. A drawback of the older method was that *all* the photographs had to be matched with *all* the traits (or character sketches) ; consequently if one mistake was made the matcher was forced to make at least one more. For instance, if he had two photographs to match with two sketches he would be either completely right or completely wrong. The more modern technique is to use more photographs than sketches in the first part of the matching, and then more sketches than photographs in the second. More reliable results are thereby obtained.

Although this section opened with an attack on the over-optimistic excesses of some of the work associated with " p " tests, it would be unfair to close it without a reference to multiple factor analysis, as it has been developed principally in the United States. The technique has been found particularly useful in putting the classification of the test results on to a mathematically reliable basis, and also in the elucidation of overlapping variables, such as, for instance, may exist in a test or questionnaire on *interests*. For a dispassionate account of the work the reader is referred to P. E. Vernon's report (No. 83) to the *Industrial Health Research Board* (H.M. Stationery Office, 1938).

4. The Study of Expressive Movements.

Allport and Vernon§ used a battery of tests of normal expressive activity. The tests consisted of speed of reading and counting aloud ; speed of walking and strolling ; length of strides ; estimation of distances between the hands ; drawing circles the estimated size of 25 cent. and 50 cent. coins, and drawing rectangles the estimated size of dollar bills ; estimation of distances outwards

* P. E. Vernon, *op. cit.*, p. 2.

† For a summary of work on matching, see P. E. Vernon, " The Matching Method Applied to Investigations of Personality ", *Psychol. Bull.*, 1936, xxxiii, pp. 149–177.

‡ P. E. Vernon, " Can the ' Total Personality ' be Studied Objectively ? ", *Character and Personality*, 1935, iv, pp. 1–10.

§ G. Allport and P. E. Vernon, *Studies in Expressive Movement.* New York, 1933.

from and inwards towards the body, and estimation of angles with rotating arm; speed, extent and alignment in arranging Binet's five cubes; strength of normal grip, together with over- or under-estimation of weights; normal speed and pressure of tapping with finger and with hand; normal speed of leg-tapping and the normal pressure of fingers on a stylus; speed, area, extent, etc., of circles, squares, and parallel lines drawn by hands and feet on paper, on a blackboard and in a sandbox; gross features of handwriting, and writing *eeee*'s in sand with the foot.

The aim of the procedure was to try to discover whether these different modes of expressive movement were consistent with one another and could be regarded as expressing a single underlying feature of the total personality. The tests had, therefore, a direct psychological implication, it being a common belief that a person's gestures and expressive movements are consistent with one another.

The results showed that there was no single speed or general motility factor common to all the tests, but that there were three somewhat broad speed factors, namely, *verbal*, *drawing* and *rhythmic*. Verbal and drawing speeds were fairly closely connected, but they were almost independent of rhythmic speed.

In conclusion, Allport and Vernon maintained that their results were inconsistent both with a theory of a general psycho-motor factor, and also with a theory that action is completely specific. Their results favoured a theory of organized psycho-motor dispositions or expressive traits.

Much work on the study of expressive movement has recently been done on handwriting. Modern graphology is not concerned with the attribution of particular psychological characteristics to minute details of handwriting, such as the length of the upward or of the downward strokes, but with handwriting as a vehicle for the expression of personality, and as a form of gesture that it is more convenient to analyse than other forms of gesture, like facial expression or gait, because it is more permanent and can thus be analysed at leisure. Modern graphologists take into account the general features of the script and their interrelations with such things as the nature of the ink and the paper, the style the person was taught, his nationality, and his speed of writing, and they claim to obtain satisfactory indications of personality characteristics in this way.

5. THE FREE ASSOCIATION METHOD.

The subject is given a list of words to each of which he has to respond as quickly as he can with the first word that comes into his head. Using the Kent-Rosanoff technique, a quantitative interpretation may be put on the responses. Kent and Rosanoff* standardized their list of stimulus words on

* G. H. Kent and A. J. Rosanoff, "A Study of Association in Insanity", *Amer. Journ. Insan.*, 1910, lxvii, pp. 37–96, 317–390.

1,000 individuals by tabulating the responses they received to each of the stimulus words. It was then possible to see with what percentage of frequency any given response word was made. When the test is used on new subjects the degree of individuality or conventionality of their responses can be estimated from the Kent-Rosanoff table, and various deductions may be made about their personality characteristics. Such, at least, is the claim, though it is well to point out that the list was standardized on Americans, and standardized in 1910. It would be advantageous if a new standardization were made in this country.

6. The Rorschach Method.

An indirect measure of personality characteristics is the Rorschach ink-blot test.* Its accurate interpretation depends to a large extent on the width of the tester's clinical experience of it use. The aim of the test is qualitative, namely, to "aid the clinician in obtaining a schematized intuition into the total personality of his subject or patient".† It consists of ten bilaterally symmetrical ink-blots which are shown one by one to the patient, who is told to describe anything that they suggest to him. The patient may look at the blots in any position and for as long as he pleases. His responses are recorded and subsequently classified. The classifications include (1) the type of stimulus selected—whether the patient reacted to the blot as a whole, to a detail, to a rare detail, etc.; (2) the determinant of the response — whether the blot, or some part of the blot, suggested an experience of felt movement, or whether colour was the sole determinant of the response; whether colour was primary and form secondary, or form primary and colour secondary; or whether form alone determined the response, and if so whether the form was good or poor; (3) the content of the response, including human and animal forms, or human and animal details, common responses and original responses. Other observations are also made, for instance, the time taken before a response is made to a blot, and the total time taken over the ten blots, the degree of card turning, perseveration in responses, verbosity and elaboration, the general attitude to the experiment, and so on. According to the relative proportions of the answers in different categories, as well as according to the absolute number in each category, an estimate is made of the patient's type—such as creative, abstract, practical, analytic, organizing. Healthy superior adults, for instance, generally give more whole, movement, and form responses than the average, but fewer colour and animal responses; the feeble-minded usually give fewer whole, movement, colour and form, but more animal responses than the average. The more intelligent individual usually also gives responses that possess a

* See, e.g., S. J. Beck, *An Introduction to the Rorschach Method*, American Orthopsychiatric Assoc., 1937, pp. 278. See also H. Rorschach, *Psychodiagnostik*, Berne: Huber, 1932, pp. 230. The blots themselves are provided in vol. ii of Rorschach's book.

† P. E. Vernon, "The Rorschach Ink-blot Test", *Brit. Journ. Psychol.*, Medical Section, 1933, xiii, p. 90.

wide range of meanings, whereas if there are too many animals, or any other single category, a degree of stereotypy may be indicated. A high productivity of answers in the blots which contain colours is associated with affective release, whereas a more inaccurate, poorer and more restricted series of responses in these cards is distinctive of neurotics ("colour shock"). The analysis of the answers also throws light on vocational aptitudes by revealing certain interests that the patient possesses. But the accurate interpretation of the various factors varies with the background of the personality in which they appear, and each element has to be interpreted in the light of all the others.

7. TYPOLOGIES.

Qualitative typological approaches to the study of personality may be seen in the work of Kretschmer, Jung, Spranger and others.

(a) *Kretschmer.*

Kretschmer's typology* has a physiological basis. He found a correspondence between certain types of bodily physique and certain types of mental disorder, and extended the concept of a correspondence between them to more normal temperamental differences. He admits that his types are ideals, and that individuals may diverge from the ideal to a greater or lesser extent ; but exactly the same thing is true in medicine. Cases of measles or chicken-pox very rarely show *all* the textbook symptoms, yet no one doubts the diagnosis on that account. Kretschmer found that patients suffering from the mental disorder known as manic-depression tended to be of pyknic bodily type, a type of physique characterized by its tendency to run to fat, its medium height, rounded figure, soft, broad face, and massive neck. He also found that patients suffering from schizophrenia tended to be of asthenic (leptosomatic) or athletic bodily type, the former being characterized by leanness and tallness, and angularity. He further maintained that in normal psychology similar relationships held good. Normal individuals may be divided into cyclothymes and schizothymes on the basis of their temperamental characteristics—the two categories corresponding with the abnormal individual's manic-depression or schizophrenia—and cyclothymes tended to be of pyknic bodily type, schizothymes of asthenic bodily type. Kretschmer gives short descriptions of a few typical cases of cyclothymes, e.g., the gay chatterbox, the quiet humourist, the silent good-tempered man, the happy enjoyers of life, and the energetic practical man ; and of schizothymes, e.g., the polite sensitive man, the world-hostile idealist, the cold masterful natures and egotists and the dried and emotionally lame.

* E. Kretschmer, *Physique and Character.* London : Kegan Paul, 1925.

https://doi.org/10.1192/bjp.84.353.1008 Published online by Cambridge University Press

Kretschmer's theory crystallizes a traditional view expressed in literature and in the beliefs of the man in the street about the temperamental characteristics associated with roundness and adiposity, on the one hand, or with leanness and angularity on the other. It is also the view of many clinical workers that the theory expresses a relationship that is frequently found in practice. On a purely statistical basis, however, the results are not overimpressive when the age-factor is eliminated. It is necessary, when testing the theory, to be sure that the group of schizophrenics is of the same average age as the group of manic-depressives; otherwise a spurious correlation may be obtained. Schizophrenia is a disorder that is more likely to occur at a relatively early age, whereas manic-depression is more likely to occur later. So if a relatively young group of schizophrenics is compared with an older group of manic-depressives, it is hardly surprising to find that the manic-depressives are better covered with flesh than the schizophrenics.

(b) *Jung.*

Jung's typology* is based on the concepts of introversion and extraversion, which are held to be innately determined characteristics. The introvert is a shy and impenetrable person whose attitude towards objects in the external world is characterized by his acting as if they were continually attempting to overpower him, and had to be frustrated. The extravert is an open and sociable person, whose attitude towards objects in the external world is characterized by his plasticity in adjusting to meet his changing circumstances. The extravert will join in the general applause given to a popular cause; the introvert will hold that merely because a cause is given temporary popularity is not sufficient reason for believing it to be a good cause. The introvert tries to select out of the things going on around him those that fit in with his own mental make-up; the extravert changes his interests in conformity with the changing events in the outside world.

In addition to these two general attitudes there are also four ways by which the attitude can express itself. These are known as thinking, feeling, sensation and intuition. For instance, the extravert thinking type collects facts for their own sake; the introvert thinking type collects them in an analytic spirit, and as evidence to prove a theory. The extravert feeling type is a follower of fashion; the introvert feeling type has hidden depths of emotion under an impassive exterior. The extravert sensation type is realistic and indulges in concrete enjoyment; the introvert sensation type takes his pleasures in his own way. The extravert intuition type is guided by practical possibilities, and is the type that makes the successful merchant, speculator or politician; the introvert intuition type chases after ideals, he is a voice crying in the wilderness, and he is the type that becomes the prophet or the mystical

* C. G. Jung, *Psychological Types*. London: Kegan Paul, 1924.

dreamer. The thinking, feeling, sensation and intuition function types do not always, however, express themselves in a pure form. Certain combinations are also possible, and thinking or feeling may be combined with either sensation or intuition.

A number of questionnaires have been devised to measure extraversion and introversion. The Neymann-Kohlstedt questionnaire* contains fifty questions which the subject is asked to consider from the point of view of whether he likes them or not, and to mark them accordingly. For instance:

"11. Accept suggestions rather than working them out for yourself: YES; NO.

"24. Speak in public: YES; NO.

"25. Confide in others: YES; NO."

The examiner is provided with a key on which the typical extravert answers are shown.

Heidbreder devised a test† of introversion, based on specific characteristics of introversion, and the subject has to mark each item with a plus sign if he possesses it and with a minus sign if he does not. For instance:

"5. Do you indulge in self-pity when things go wrong?

"8. Are you critical of others?

"18. Are you radical? Do you want to change the world rather than adjust yourself to it?"

(c) *Spranger.*

Spranger's types‡ are based on attitudes. The six attitudes are theoretic, economic, æsthetic, social, political and religious.

The *theoretic* attitude is characterized by objectivity. It leads to analysis and synthesis, to reasoning and to systematization. Things are not recognized as beautiful or ugly, useful or useless, good or bad, but only as true or false. A man with this attitude is mainly interested in the discovery of truth. In his social relations he is an individualist, for his objective mental attitude does not find sympathetic support from other people, and he tends to regard family ties as relatively unimportant.

The *economic* attitude is guided by utility. A man with this attitude economizes goods and forces, time and space, in order to gain the maximum useful effect for himself. He is content to let beautiful landscapes be destroyed for economic motives, for he is interested in what is useful rather than what is beautiful. He is thoroughly practical. He is also unsocial. He is uncharitable and lacking in altruism, except in so far as his economic position may be improved by it.

* C. A. Neymann and K. D. Kohlstedt, "A New Diagnostic Test for Introversion-Extraversion", *Journ. Abn. and Soc. Psychol.*, 1929, xxiii, pp. 483–4.

† E. Heidbreder, "Measuring Introversion and Extraversion", *ibid.*, 1926, xxi, pp. 120–34.

‡ E. Spranger, *Types of Men*. Halle: Niemeyer, 1928.

The man with the *æsthetic* attitude is unambitious. The highest good is to be found in form and harmony. Knowledge is no good for its own sake, but only in its relation to other pieces of knowledge which fit together harmoniously. To describe an object in terms of its utility or practical value destroys its beauty. The æsthetic man is not unsocial, but he tends to be eccentric, and since he introduces æsthetic values into his social relationships he is apt to be difficult and intolerant.

The *social* attitude is guided by living through and for others. It is in sharp contrast with the economic attitude, and it also contrasts with the æsthetic attitude. A man with a predominantly æsthetic attitude, for example, avoids people who live in poverty and squalor, whereas the man with a predominantly social attitude feels that the greater the degree of poverty and squalor the more necessary is it to give help, and the more he seeks out such people.

The *political* attitude is characterized by power. Knowledge is useful only in so far as it leads to the gaining of power over other people. Power may be acquired by tact, or by force, or by any other available method. A man with the political attitude tends in his social relationships to mix with people so that he may dominate them.

The *religious* attitude directs the individual towards the highest possible values of an integrated life. An intellectual system as such implies nothing in regard to total value. The religious attitude has a strong positive social relationship, although in certain cases the religious man may try to free himself entirely from the world and live the life of a hermit.

For purposes of description the types are isolated, but Spranger admits that each attitude may be found in varying degrees in all personalities. In general, there is a correspondence between economic and political values, between social and religious, and between æsthetic and theoretical. On the other hand, both social and religious values are opposed to theoretical; and economic and political values are opposed to æsthetic and to religious values.

A questionnaire based on Spranger's types has been devised by Allport and Vernon.* The object of the questionnaire is to avoid the isolation and measurement of single personality traits, and to study the personality as it expresses itself as a whole in Spranger's six levels of value, so as to get a picture of the relative predominance of the six values in the personality as a whole. The situations described in the test are made as natural as possible, so that the subject may picture himself in the situation without artificiality, and arrange the alternative answers in the order of his preference accordingly. The following are two examples of such situations:

> "The main object of scientific research should be the discovery of pure truth rather than its practical applications. (*a*) Yes; (*b*) No."

* P. E. Vernon and G. Allport, "A Test for Personal Values", *Journ. Abn. and Soc. Psychol.*, 1931, xxvi, p. 3.

"Do you think that a good government should aim chiefly at:

"(*a*) More aid for the poor, sick and old.

"(*b*) The development of manufacturing and trade.

"(*c*) Introducing more ethical principles into its policies and diplomacy.

"(*d*) Establishing a position of prestige and respect among nations."

The questionnaire is claimed to have high reliability and validity, and to be useful for such purposes as vocational guidance.

8. THE INTERVIEW.

The method that in practical life is used far more than any other method so far described is the interview. Doctors and anthropologists, solicitors and social workers, employers, vocational advisers and psychiatrists, together with numbers of other groups of people, make use of the interview as part of their stock-in-trade. For making a preliminary survey of a field, however, the interview is wasteful compared with the questionnaire, for results must be gathered individually instead of in bulk. It is also in one respect a less reliable method, for suggestion may be more easily conveyed by a particular tone of voice, emphasis and rhythm, than by the chilly detachment of the printed word. It has, however, the advantage over the questionnaire that if misconceptions arise they may be more easily dispelled. If someone misunderstands a question in a questionnaire his answer must be accepted at its face value, but if he misunderstands a question in an interview, further discussion may clarify the issue. In one investigation, for example, it was found that a certain group of textile workers understood the word "arbitration" to mean "surrender". The question, "Are you in favour of arbitration", therefore, meant, "Are you in favour of surrendering to the employers", and it was greeted with a universal negative. In a subsequent interview, the true meaning of the word was demonstrated by example and illustration, and a different pattern of response to the question was obtained.

The technique of conducting an interview varies with the interviewer, the person interviewed and the purpose of the interview. Some have advised a very rigid standardization of the interview situation, going even to the length of eliminating oral questions and presenting the person interviewed with a card bearing typewritten questions. The assumption presumably is that an objective standardization of the conditions presents each person with an equivalent situation. This is a very dubious assumption, as we have seen. Different individuals with different personalities may react differently to the same objective situation. Such a degree of standardization combines the disadvantages of the questionnaire method with none of its advantages. Most interviewers, however, believe in making the interview situation flexible.

The first thing of importance is to gain *rapport*, and the skill which the interviewer employs to effect good *rapport* depends partly on his natural skill in handling people, but, in addition, very largely on the acquisition of technique, based on a wide clinical experience. The solicitor and the doctor and the vocational adviser have this advantage over other groups of interviewers, that the patient comes to them with a definite problem on which he needs advice. He is thereby more ready to meet the interviewer half way. But the psychologist interviewing a child who has been brought unwillingly by his parents to a clinic often has a more difficult job. He must see to it that the child's suspicion and antagonism are removed, and that he is willing to accept the psychologist as someone on his side, before the serious purpose of the interview is undertaken. If the psychologist can gain from the case-history or from the parent some information about the child's interests or general attitudes before he talks to the child himself, he may facilitate his attempt to win his confidence. If possible, matters should be discussed first of all which are likely to afford the person interviewed a certain amount of pleasure. Failing these, the topics should be at least neutral in emotional content. Matters which are unpleasant should always be left until later, and when disposed of, the interview should be finished by referring once more to a pleasant topic.*

As a general rule the interviewer should never, of course, criticize or moralize over the information he receives in the course of the interview. Although the facts he elicits can sometimes be accepted at their face value, yet the attitudes they betray are often important. If, for example,† a social worker discovers that a mother continually refers to one child as a little angel and to another as a little devil, it is not only necessary to check up her opinion of the two children by questioning neighbours, school teachers, the father, and so on, but it is also important to appreciate the effect of the mother's attitude towards the two children on the home situation.

Many interviewers make use of some of the other techniques that have been described in this chapter. Thus, rating scales are sometimes found to be useful for purposes of quantifying some aspects of a person's attitude or behaviour. Descriptive adjectives are usually avoided, and in their place are given actual descriptions of behaviour. Thus, for example, it is obviously more valuable to report that a mother continually referred to one child as a little angel and to the other as a little devil than to report that the mother showed a favourable attitude towards one child and an antagonistic attitude towards the other. Adjectives like "restless", "energetic", "lazy" and so on should be replaced by the report of an actual incident which illustrates such behaviour. The results of specific tests, too, may be more illuminating than descriptions in broad categories like "poorness of memory".

* See A. Rodger, *Problems and Methods in the Study of Society* (ed. F. C. Bartlett *et al.*). London: Kegan Paul, 1939.

† S. Clement Brown, *Problems and Methods in the Study of Society* (ed. F. C. Bartlett *et al.*). London: Kegan Paul, 1939.

The skilled interviewer is always on his guard against the different meanings that different people attach to the same word, and experiments have demonstrated how fallible interviews may be, and the bias that may be conveyed in interviews from the interviewer to the interviewed. Rice,* for example, analysed the records of the interviews by twelve investigators of 2,000 homeless men, who were applying for lodging. One of the interviewers was an ardent prohibitionist and found that the downfall of 62% of the applicants was due chiefly to liquor, while only 7% of the applicants had been seriously affected by industrial conditions. Another interviewer, who was a socialist, found that only 22% owed their condition to liquor, and that 39% had been seriously affected by industrial conditions. Still more enlightening was the fact that the prohibitionist said that 34% of the applicants had *themselves* given liquor as the cause and 42·5% industrial conditions, whilst the socialist said that only 11% blamed liquor and 60% blamed industrial conditions.

The careful interviewer does well to remember the warning implied in Bertrand Russell's description of the typical American and the typical German approaches to the study of animal behaviour :

"One may broadly say that all the animals that have been carefully observed have behaved so as to confirm the philosophy in which the observer believed before his observations began. Nay, more, they have all displayed the national characteristics of the observer. Animals studied by Americans rush about frantically with an incredible display of hustle and pep, and at last achieve the desired result by chance. Animals observed by Germans sit still and think and at last evolve the solution out of their inner consciousness."†

The more experienced the interviewer the more he needs to guard himself against generalizations and descriptions that fit in with his previous experience. The greater the clinical experience the easier is it to acquire a belief in stereotypes. The more often one has observed that characteristics A, B, C and D tend to be associated with characteristics W, X, Y and Z, the more is one likely to conclude that all individuals with characteristics A, B, C and D must *ipso facto* possess characteristics W, X, Y and Z. The influence of such false stereotypy has been shown, for example, in a comparison between composite photographs of artists' sketches of criminals and composite photographs of criminals themselves. The sketches are endowed with "criminal type" characteristics which are not found in the actual criminals.

Furthermore, everyone likes to think of himself as a skilled interviewer, and to console himself with the thought that however unreliable other people may be, yet his own results are above reproach. The skilled interviewer is aware of this human defect, and guards himself against it accordingly, to the best of his ability. There is little doubt that the interview is a very important

* S. A. Rice, "Contagious Bias in the Interview", *Amer. Journ. Soc.*, 1929, xxxv, p. 420.

† B. Russell, *An Outline of Philosophy*. London, 1927, pp. 32–3. Cited by P. E. Vernon, *Brit. Journ. Psychol.*, 1933, xxiv, p. 156.

method of obtaining information about personality characteristics. It is particularly important when used to check material that has been obtained by some other means, such as by test or questionnaire, or rating scale. It is, however, essential to remember that it is a clinical method, and that its results are more often qualitative than quantitative. For this reason particular care and enlightened criticism must be employed when evaluating its results.

9. The Case History.

In the case-history method much of the information about an individual is obtained through interviews, and during the interviews the precautions already described must be employed. Much of the material is, however, of a factual type, and to this extent may be more readily accepted than information bearing upon attitudes and feelings. Case-histories are obtained, too, by checking the information by referring to more than one source. Thus information received from the patient himself may be checked by referring to his relatives, his professional record, his near associates and so on. Due regard is also paid to other factors that may have influenced his present personality, for instance, to his early development, to neurotic traits, like bed-wetting or nail-biting, to the strength and development of his sex interest, to his adaptations at school, and to any special aptitudes he may have exhibited there, to his social adaptation and present sex behaviour, and so on, anything, in fact, which may be useful in interpreting his present personality. The present techniques, which are adopted widely in hospitals, have been evolved more or less by trial and error. It is likely that experimental workers will turn their attention to the method with a view to discovering any further refinements of technique.

III. Some Applications of Tests.

The two preceding sections concerned themselves with the principle methods of measuring personality characteristics, and in the course of the discussion some lines of application emerged. The emphasis, however, was laid on the methods rather than on the applications. In the present section the balance will be reversed.

1. The Application of Intelligence Tests to Adults of Different Ages.

It was remarked in Part I that intelligence tests could often be less appropriately applied to adults than to children, partly because adults were sometimes apt to behave antagonistically rather than co-operatively. An additional, or perhaps contributory, factor to this antagonism is the fact that children are accustomed to receiving tests and examinations of different kinds, whereas most adults discard them with their school days. This fact enhances the difficulty of interpreting with any exactitude the meaning of the scores which

adults obtain in an intelligence test. It is ridiculous, for example, to regard an adult with a mental age of IX years as identical with a normal child of 9. Mental ages applied to adults cease to possess the practical meaning that is clearly theirs when they are used as a basis for comparing two children. The only line of defence is to argue that an adult with a mental age of IX has more in common intellectually with an average child of 9 than he has with a normal adult. At the same time Simon,* who suggests that this is so, says that he prefers to use more qualitative terms like moron, imbecile, idiot, etc., comparable to terms like " dwarf " for the physically defective, when he is referring to the intelligence of subnormal adults. It would appear that a true basis for comparison between adults would emerge only from tests standardized on adults themselves. Possibly it would ideally be desirable for completeness to standardize fresh tests on different groups of adults, e.g., morons, mental defectives, normals, superior adults, and so on, so that an individual could be compared, if it was thought desirable, with more than one of the different groups. As, however, there is little likelihood of the proposal being adopted in the near future, a more practical proposal might be to try out the technique of testing by batteries instead of by scales† until a sufficient number and range of independently standardized tests had been developed to allow scope for the selection of those that were most appropriate to adults of different types.

When existing tests have been applied to adults of different ages, a decline in efficiency with age has often been reported. For instance, Jones and Conrad‡ gave an intelligence test to nearly 1,200 people varying in age from 10 to 60. They found a rapid increase in intelligence between 10 and 16 years, a slower increase to between 18 and 21 years, and then a decrease to 55 years, which was equivalent to the average at the age of 14. They were of the opinion that the decrease was a genuine one, and was not attributable to poorer motivation, hearing, sight, speed or similar factors.

Miles,§ however, found that the principal decline with increasing years was in physiological abilities, and that abilities on the intellectual level showed a loss more in speed of response than in accuracy. He investigated the scores of 800 men and 800 women, varying in age from 10 to 90 years who were given individually a large number of different tests in four consecutive half-hour sessions.

In interpreting the result of tests on adults it is, however, as was remarked before, necessary to take into consideration the effect of rustiness due to lack of practice. Sorensen‖ is of the opinion that this factor is sufficient to account

* T. Simon, *Ann. Med.-Psychol.*, 1936, xciv, pp. 474–477, 477–480.

† See Part I, Section 4.

‡ H. E. Jones and H. B. Conrad, " The Growth and Decline of Intelligence ", *Genet. Psychol. Monog.*, 1933, xiii, pp. 233–298.

§ W. R. Miles, " Age and Human Ability ", *Psychol. Rev.*, 1933, xl, pp. 99–123.

‖ H. Sorensen, " Adult Ages as a Factor in Learning ", *Journ. Educ. Psychol.*, 1930, xxi, pp. 451–459.

for the decline. He found a negative correlation between age and learning ability in a group of adults who had not studied for a long time, but no such correlation for two other groups of adults who had continued learning. He therefore concludes that there is no true decline in ability to learn with age. Rustiness, no doubt, partly accounts for the decline, but it is doubtful whether it is of sufficient importance to account for it all.

2. Deterioration.

Within the last few years an increasing interest has been shown in the use of tests for differentiating between mental deterioration and mental defect. Hart and Spearman,* as long ago as 1914, gave different groups of psychotic patients a large number of tests and compared their performance with that of normal people. But most of the recent interest in the subject followed the publication of the Babcock deterioration test† in 1930. The theory on which the test is based is that an individual's score in the vocabulary test is little affected by deterioration, whereas his score in tests, which involve the application of new knowledge, may be greatly affected. When, therefore, there is found to be a large discrepancy between his scores in the two types of test, the vocabulary score being, let us say, that of the normal adult, while his average score on the other tests is that of the normal child of 8, deterioration is indicated; whereas if the adult obtains a score of the normal child of 8 on both the vocabulary and the other tests, original defect is indicated. Babcock presented her tests as a battery, each test being standardized independently for all mental ages. It is not, therefore, necessary to use every test in the battery when it is applied. Subsequent work has confirmed the belief in the validity of the vocabulary score as an indicator of original mental level, except in very severe cases of mental deterioration, or in some cases of general paralysis, and Wittman‡ has attempted to show that different types of organic deterioration, e.g., epilepsy, paresis, arteriosclerosis, and alcoholism, give different patterns of response in the Babcock deterioration test; some types doing relatively better or relatively worse than others in different test items.

The work of Hart and Spearman was continued by Simmins,§ who applied visual-perceptual and vocabulary tests to 200 mental hospital patients, compared the patients' scores on the two types of test, and so obtained a measure of the degree of their deterioration.

* B. Hart and C. Spearman, "Mental Tests of Dementia", *Journ. Abn. Psychol.*, 1914, viii, pp. 217–264.

† Babcock, "An Experiment in the Measurement of Mental Deterioration," *Arch. of Psych.*, 1930, No. 117, pp. 105.

‡ P. Wittman, "The Babcock Dererioration Test in State Hospital Practice", *Journ. Abn. and Soc. Psychol.*, 1933, xxviii, p. 70.

§ C. Simmins, "Deterioration of "g" in Psychotic Patients," *Journ. Ment. Sci.*, 1933, lxxix, pp. 704–734.

The same problem is approached by Malamud and Palmer,* who use, not the Babcock test, but the shortened form of the Stanford-Binet test. They compared the performance of 100 organically deteriorated patients, 100 schizophrenics and 100 feeble-minded patients, all of whom had a mental age of between 8 and 12 years. They showed that patients with organic deterioration found the vocabulary test and the president and king test easier, but counting backwards, digits backwards, the date, weights, three words, absurdities, and associations tests harder than the feeble-minded found them. The schizophrenics found the vocabulary test, the president and king test, and arithmetical reasoning easier, but comprehension, date, weights, three words, absurdities, associations, pictures and problems tests more difficult than the feeble-minded patients found them. Malamud and Palmer maintain that these differences are sufficiently marked to be described in terms of objectively recognizable patterns of scatter, which are useful both in differentiating diagnostically between the three groups and also for purposes of gaining insight into the qualitative nature of deterioration.

A similar attempt to obtain an objectively recognizable pattern of response in the Stanford-Binet test from epileptics has been made by Collins, Atwell and Moore.†

3. Diagnosis and Prognosis from "Scattergrams".

"Scatter" in test results, that is to say the distribution of successes and failures over a wide range of mental ages, has often been regarded as significant of maladjustment, and Earl‡ has recently suggested a method for diagnosing different types of emotional disorder among feeble-minded adults. He gives his patients the vocabulary test, an oral absurdities test, and two performance tests (Kohs' blocks and the Danvers-Dearborn form-board), and expresses the results of each of them, together with the means of the two verbal tests, the means of the two performance tests, and the mean of all four tests, as percentiles, presenting them in the form of a graph. Such "scattergrams" have several typical forms: The percentile ranking of the patient may be approximately the same in all the tests, giving a normal picture; or the scattergram may be verbally biased, that is to say, the percentile ranking of the patient in the verbal tests may be considerably higher than in the performance tests; or it may be biased on the performance side. In a verbally biased scattergram the percentile ranking of the patient may be approximately the same, or differ, in the two verbal tests. If it is biased on the performance side there are three corresponding types of curve. Earl claims that such differences in scattergrams provide valuable aids to diagnosis and prognosis. For example,

* W. Malamud and E. M. Palmer, "Intellectual Deterioration in the Psychoses", *Arch. Neur. Psychiat.*, 1938, xxxix, pp. 68–82.

† A. L. Collins, C. R. Atwell and M. Moore, "Stanford-Binet Response Patterns in Epileptics", *Amer. Journ. Orthopsychiat.*, 1938, viii, pp. 51–63.

‡ C. J. C. Earl. (At present unpublished.)

if the scattergram is markedly biased on the performance side and at the same time the vocabulary result is below the absurdities result, an "averbal" group of a mildly excitable type is indicated. If the scattergram is verbally biased and at the same time the Dearborn result is definitely below the Kohs' result, there is an indication of a fairly general unreliability of inhibition, or else of gross emotional disorder, with grave prognosis. In general, any marked scatter of results seems to be significant of emotional maladjustment. Earl, however, insists that these scattergrams must be interpreted very cautiously, and always in the light of the quality of behaviour shown during the test.

4. Undiscovered Superiority.

In schools the application of intelligence tests has sometimes led to the discovery of children whose mental superiority was previously unsuspected. A child of superior intelligence, if his superiority has not been recognized, may find insufficient means of expression in a class of children whose average intelligence is much below his own. Consequently he may exhibit unsocial behaviour and pay little attention to his school work, with the result that he does not appear superior, but rather inferior, to the rest of the children. The disclosure of his superior intelligence makes possible special treatment, directed towards enabling him to find more adequate outlets.

5. Educational Retardation.

In other cases the application of an intelligence test may indicate that a child who is backward in school subjects is yet of normal intelligence. Sometimes the backwardness is general, but more often it is specific to reading or arithmetic. Educational backwardness may be discovered by applying standardized tests for different school subjects, e.g., reading, arithmetic, spelling and comparing the child's educational age in the subjects with his mental age. This provides what Burt calls an "achievement ratio" $= \frac{\text{educational age}}{\text{mental age}} \times 100$. The achievement ratio may be calculated for each school subject separately or for all together. If the ratio is 100, then the child's scholastic attainments are keeping pace with his innate intelligence. If the ratio is much below 100, it may indicate that the child has missed a lot of schooling, or that he is lazy, or that unsuitable methods of teaching have been employed. The last may be remedied by providing him with special, individual coaching by someone who has specialized in the work. Such children form the most satisfactory material for special coaching. Far less hopeful results are obtained when special coaching is given to educationally retarded children, whose level of intelligence is also much below normal.*

* For a full discussion of the problems connected with both educational retardation and generally backward children, see C. Burt, *The Backward Child*. London: University of London Press, 1937.

6. Intelligence and Delinquency.

The question of the relationship between intelligence and delinquency is one that is often raised, but the evidence is somewhat conflicting.* Some studies show that delinquents, particularly those in institutions or those who come before the courts, are, on the average, inferior in intelligence to the general population, while others suggest that they compare favourably with norms taken from among individuals in the same socio-economic grade as the delinquents. It must be remembered, too, that delinquents of superior intelligence are less likely to be caught than those of inferior intelligence, with the result that the average level of intelligence of the whole group of delinquents is probably to some extent above the level indicated by tests applied to those available for investigation.

7. Vocational Guidance.†

In vocational guidance an individual is examined in order to inform him what job or jobs he is best fitted to do. The appointment of careers masters in schools is an indication of the increasing importance that is attached to the work. Often the recommendations have to be made on the result of a single interview, but the National Institute of Industrial Psychology point out that more effective work could probably be done if the recommendations were based on a series of observations extending over a period of years. Such a procedure would necessitate a more widespread use of careers masters than exists at present. However, even with the present limitations, considerable success has already been achieved.

The individual to be advised is given a series of tests. His intelligence is tested, both by verbal and by non-verbal tests; he is given tests for special abilities, such as manual and mechanical, constructive, artistic and æsthetic, and he is given ratings on temperamental characteristics. A report is obtained on his home conditions—the type of home he lives in—on his father's occupation (and that of his mother also if she has one); on his parents' ideas about his occupation and what they think of his capacity for it; and on their description of his character, interests and behaviour. A report is obtained from the school medical officer about his physical condition, to see if he has any special ailments, such as anæmia, asthma, tuberculosis and so on—all of which are of obvious importance before the selection of an occupation is made. And finally the individual is himself consulted and interviewed with a view to finding out what are his main interests, wishes, etc., and whether he has any special type of emotional maladjustment. His more general temperamental characteristics will have revealed themselves to a large extent in the test

* For a short summary of the evidence and for further references see C. M. Loutit, *Clinical Psychology*. New York: Harpers, 1936, pp. 373–375.

† See, e.g., A. Macrae, *Talents and Temperaments*. London: Nisbet, 1932. Also C. A. Oakley and A. Macrae, *Handbook of Vocational Guidance*. London: University of London Press, 1937.

situations by the way he reacted to them, e.g., whether he reacted in an inhibited or in an aggressive way; whether his movements were fluent, or sporadic and awkward; whether he was quick or slow; whether he was co-operative or negativistic; whether he was garrulous, articulate and possessed a wide vocabulary, or inhibited, inarticulate and limited in his vocabulary; whether he was self-conscious or not; whether he was affected by praise, either advantageously or deleteriously, or unaffected by it; how he reacted to being made to hurry—whether he made more mistakes and increased his time and his ability to co-ordinate his movements, or remained relatively unaffected or improved by it; and so on. In this way a fairly complete picture of his personality can be built up.

In the next place an analysis of possible occupations is prepared. The analysis often differs in different parts of the country, according to the different kinds of jobs that are available there. In a broad way, however, it is possible to classify occupations into the following categories:

(1) *Higher professional and administrative.*—Lawyer, doctor, teacher, scientist, civil servant, etc.

(2) *Lower professional.*—Elementary teacher, secretary, reporter, surveyor, merchant.

(3) *Clerical and highly skilled.*—Shorthand typist, office clerk, telegraphist, nurse, compositor, machine inspector.

(4) *Skilled work.*—Tailor, engine driver, policeman, mechanic, fitter, plumber, carpenter, farmer, shop assistant.

(5) *Semi-skilled.*—Mechanical repetitive work, poorer commercial, domestic servant, miner, carter, baker.

(6) *Unskilled.*—Repetition work, manual labour, automatic machine worker, porter, packer, etc.

These groups roughly correspond to the levels of intelligence that is required, although there is considerable overlap. For example, Cattell* found the following median intelligence quotients in the undermentioned occupational groups:

30	Secondary school teachers	151
25	Physicians and surgeons	146·5
50	Central school teachers	145
20	Civil engineers	142
18	Mechanical engineers	140
90	Elementary school teachers	137
15	General managers in business	137
57	Shorthand typists	129
54	Commercial clerks	127
14	Typists	126
24	Commercial travellers	123

* R. B. Cattell, "Occupational Norms of Intelligence", *Brit. Journ. Psychol.*, 1934, xxv, p. 1.

250	Nurses (probationers)	122
52	Precision fitters	114
14	Coach-body builders	106
20	Sheet metal workers	102
12	Shop assistants	99
33	Carpenters	98
31	Cabinet makers	97
33	Machine operators	96
24	Coach trimmers	96
12	Hairdressers	89
16	Upholsterers	87
12	Welders	87
19	Factory packers and sorters	78

Occupations can then be reclassified according as they require:

(*a*) Special physical qualifications, or the absence of certain defects, e.g., colour blindness, deafness, delicate touch and so on.

(*b*) Verbal or linguistic ability.

(*c*) Mechanical and practical ability.

(*d*) Social ability.

(*e*) Artistic ability.

(*f*) Special educational attainments, e.g., writing, spelling, calculating.

(*g*) Special character or temperamental qualities, e.g., honesty, truthfulness, reliability, patience, tactfulness, initiative, and so on.

Having classified the occupations, and having obtained a picture of the individual who is to be advised, the next step is to fit the two together. The recommendations, when they are finally given, are made on rational grounds, and no attempt is made to *force* a person into a particular job. The recommendations are usually wide enough to allow him to make a choice between two or three occupations within the same group if he wants to, so that if a particular job is unobtainable he will have alternative choices. Here the importance of working in close co-operation with the employment exchanges becomes obvious. The vocational adviser must, in addition to everything else, discover the present and the probable future demand for labour in different occupations before he makes his final recommendations.

Various investigations have been made to discover how far vocational guidance has been successful,* and, on the whole, the results have been very satisfactory. Various groups have been compared, e.g., those who have been vocationally advised and have entered jobs in accordance with the recommendations; those who have been advised and have entered jobs not in

* See, e.g., C. Burt and others, "A Study of Vocational Guidance", *Industrial Fatigue Research Board Report*, 1926, No. 33. F. M. Earle and others, *Methods of Choosing a Career*, London, 1931. E. P. Allen and P. Smith, *The Value of Vocational Tests as Aids to Choice of Employment*, Education Committee, Birmingham, 1932.

accordance with the recommendation ; and control groups who have entered the labour market in the ordinary way. On the whole it has been found that those who have taken jobs in accordance with the recommendations are, when they are followed up some years later, making more money, giving more satisfaction to their employers, changing their jobs less frequently, happier in their jobs, etc., than the individuals in any of the other groups. The advice has not been completely successful in every individual case, but that may be largely due to the need for a further improvement in the technique of advice, e.g., improved methods of estimating temperamental characteristics, and longer period observations before the recommendation is made.

In a recent experiment in vocational guidance in Borstal institutions, Rodger* found that of 158 boys placed in work parties recommended by the adviser 69·5% became grade A workers, whereas of the 160 boys in the control group who were placed in work parties recommended by their house-masters only 45·6% became grade A workers.

8. Vocational Selection.

In vocational guidance, as we have seen, the object is to find an appropriate job for a particular individual. In vocational selection the approach is from the opposite direction. Here the object is to find an appropriate individual to fit a particular job. The employer wants to choose the most suitable workers for his job. One method of selection, and perhaps the commonest, is the *interview*, which we discussed in the last chapter. But another is to use tests.

The tests used in vocational selection may be conveniently classified into four types.

(1) *Sample tests.*—In this type of test the individual simply does a sample of the work for which workers are required. For example, the typist might be tested on a typewriter, or the chauffeur on a car. The trouble about this kind of test is that it shows *proficiency* rather than *aptitude*. It tests how good the individual is now, ignoring the fact that different applicants may have had different amounts of practice. It gives no indication of future performance. Yet since the actual work to be done by the applicants is sometimes quite different from what they have done before, what is really required is some estimate of future rather than of present performance. It is for this reason that in scientific vocational selection one of the other three types of test is usually employed.

(2) *Analogous tests.*—In this kind of test something psychologically similar to the work to be done is used. For example, in a test for motor drivers the subject has to react on a dummy control and his reactions are recorded. In one investigation it was found that those who failed on the test averaged three

* A. Rodger, " A Borstal Experiment in Vocational Guidance ", *Industrial Health Research Board, Report No.* 78. London : H.M. Stationery Office, 1937.

accidents per man over a given period, whereas those who passed had only 1·3 accidents per man. Furthermore, 46·1% of those who had no errors on the test had no accidents during the period, while 18·8% of those who had one to three errors on the test had no accidents, and only 12·5% of those having four or more errors on the test had no accidents.

(3) *Analytic tests.*—In this type of test the job is analysed into its independent parts, and each of these parts is tested separately. Thus, in analytic tests for shooting, Banister analysed the abilities necessary for good shooting into a number of different parts (for example, steadiness, aiming, constant pressure on the trigger, and so on) and devised tests for these factors separately. One danger of the method is the fallacy known in logic as " composition ". It cannot always be correctly inferred that a person who does well on tests for separate qualities will also do well when the separate qualities are combined together in the total operation. The quality of the separate operations may undergo some vital change when they are combined together.

(4) *Empirical tests.*—In these tests there is no obvious reason why a particular test is a good measure of a job, but it has been found in practice to be so. Thus, for example, there is no immediately obvious reason why the ability to follow an argument under conditions of diminished oxygen pressure should be related to the ability to make a good aeroplane pilot, yet such has been found to be the case.

In effect, of course, all tests are empirical in the last resort, because if a test is found to be non-diagnostic for a particular ability it will have to be discarded. Yet, in the case of the tests grouped under the heading of " empirical " there is, as was said before, no obvious psychological reason why they should be good tests, whereas in the analytic and analogous tests there is.

For validating tests two methods have been employed. One method is to give the tests to apprentices and to follow their subsequent careers to see whether the results of the tests coincide with the ability of the apprentices when they become more skilled at the job. But this method takes time before the diagnostic abilities of the test are made apparent. The other method is to apply the tests to workers who are already doing a job and then to correlate their scores with their proficiency, or with the wages they are paid—if they are paid by piecework. Then the tests which correlate most highly are retained. But this method has the disadvantage that proficiency may sometimes depend on the degree of practice rather than on innate ability.

9. TESTS FOR ACCIDENT PRONENESS.*

Apart from external factors which give rise to accidents, such as poor lighting, heating, ventilation and so on, there is another factor of great

* See, e.g., E. Farmer and E. G. Chambers, " A Study of the Personal Qualities in Accident Proneness and Proficiency ", *Industrial Fatigue Research Board, Report No.* 5. London: H.M. Stationery Office, 1929. Also E. Farmer, E. G. Chambers and F. J. Kirk, " Tests for Accident Proneness ", *ibid., Report No.* 68. London: H.M. Stationery Office, 1933.

importance. This is known as "accident-proneness". Much experimental work has been devoted to studying it. It has been found that the frequency of accidents is not equally distributed among individuals, but that some individuals are naturally more liable to have accidents than others. In one investigation* it was found that 16% of 200 women munition workers had 46% of the accidents, and that there was a high correlation between different kinds of accidents, between accidents during different periods and between accidents at home and in the factory.

The nature of accident proneness has been investigated and has been found to correlate with certain tests known as "æstheto-kinetic" tests. Such tests are dotting, the pursuit meter (in which the individual has to follow as exactly as possible an irregularly moving pointer), reaction times, steadiness and so on. In one investigation in which the tests were used the accident-rate of the worst 25% of a group was about two and a half times that of the remaining 75%.

10. The Application of Tests to Different Races.†

Early in the history of intelligence tests attempts were made to measure the relative intelligence of different races, and the results were mostly detrimental to all the groups other than the one on whom the test had originally been standardized. But before such results can be accepted, a large number of difficulties must be eliminated, of which we will mention only two.

Still more than when one tests the intelligence of adults by tests that have been standardized on children, does the difficulty arise that different cultural experiences may influence test results. In the second place, the kind of abilities measured by intelligence tests may have different social significance in one society from that in another. Success in other societies, particularly primitive societies, often depends on quite other abilities than those measured by intelligence tests. Even performance tests of intelligence are not always suitable in every culture.

At the moment, the question of the relative intelligence of different groups has little meaning. One suggestion‡ is that a test particularly appropriate to group A should be standardized on group A, another particularly appropriate to group B should be standardized on group B, and so on. Then group A's test should be given to group B, and group B's to group A. In that way one would be able to compare the relative ability of the two groups in performing the other's test, but such tests have still to be devised and standardized.

* M. Greenwood and H. M. Woods, "The Incidence of Industrial Accidents, with Special Reference to Multiple Accidents", *Industrial Fatigue Research Board, Report No.* 4. London: H.M. Stationery Office, 1919.

† See, e.g., S. F. Nadel, *Problems and Methods in the Study of Society* (ed. F. C. Bartlett *et al.*). London: Kegan Paul, 1939.

‡ G. Schwesinger, *Heredity and Environment.* New York: Macmillan, 1933, pp. 42–3.

GENERAL REFERENCES.

ALLPORT, G. W.—*Personality: A Psychological Interpretation.* New York: Holt, 1937.
ARTHUR, G.—*A Point Scale of Performance Tests.* New York: Commonwealth Fund, 1930.
BURT, C.—*Mental and Scholastic Tests.* Second edition. London: P. S. King, 1922.
Idem.—*The Backward Child.* London: University of London Press, 1937.
CATTELL, R. B.—*A Guide to Mental Testing.* London: University of London Press, 1936.
FRYER, D.—*The Measurement of Interest.* New York: Holt, 1931.
GESELL, A.—*The Mental Growth of the Pre-School Child.* New York: Macmillan, 1925.
GOODENOUGH, F. L., and ANDERSON, L. E.—*Experimental Child Study.* New York: Century, 1931.
GUILFORD, J. P.—*Psychometric Methods.* New York: McGraw Hill, 1936.
HAMLEY, H. R.—*The Testing of Intelligence.* London: Evans, 1935.
KNIGHT, R.—*Intelligence and Intelligence Tests.* London: Methuen, 1933.
PINTNER, R., and PATERSON, R.—*A Scale of Performance Tests.* New York: Appleton, 1917.
STUTSMAN, R.—*The Mental Measurement of Pre-School Children.* New York: World Book Co., 1931.
SYMONDS, P. M.—*Diagnosing Personality and Conduct.* New York: Century, 1931.
TERMAN, L. M.—*The Measurement of Intelligence.* Palo Alto: Stanford University Press, 1916.
Idem and MERRILL, M. A.—*Measuring Intelligence.* London: Harrap, 1937.
VERNON, P. E.—*Medical Research Council: Industrial Health Research Board, Report No.* 83. London: H.M. Stationery Office, 1938.