AN EXPERIMENTAL APPROACH TO DIAGNOSTIC PSYCHOLOGICAL TESTING.

By M. B. SHAPIRO, M.A.,
Department of Psychology, Institute of Psychiatry, Maudsley Hospital, London.

OUTLINE.

A. Introduction.
B. Function and Characteristics of a Test.
C. Difficulties Arising in the Use of Tests.
    1. The Problem of Error.
    2. Inadequacy of Information.
    3. Difficulty and Expense of Production.
    4. Inherent Limitations.
D. Various Methods of Using Tests.
    1. The "Qualitative" Approach.
    2. The "Flexible" Approach.
    3. Experimental Method.
E. An Example of the Application of Experimental Method.
F. How the Experimental Method Helps to Meet the Main Difficulties.
    1. The Lessening of Error.
    2. Providing an Alternative to the Standardized Validated Test.
    3. Overcoming the Inherent Limitations.
G. Outcome of the Experimental Method.
    1. Testing only to Answer Definite Questions.
    2. Limiting "Test Interpretation."
    3. The Abandonment of the "Artistic" Approach.
    4. Qualifications of the Goldstein-Scheerer Approach.
H. Obstacles in the Way of Developing Experimental Procedures.
    1. Pressure from Colleagues.
    2. Understaffing.
    3. Misconceptions.
        (*a*) The Role of the Psychologist.
        (*b*) Multiple Causation.
I. Discussion.
J. Concluding Summary.

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

A. Introduction.

The symposium of the 1950 General Meeting of the British Psychological Society on the use of tests in clinical psychology showed that there was considerable difference of opinion on the way in which tests could and should be used in diagnostic work. As testing is one of the main activities of psychologists working in the clinical field, it is of some importance that this discussion be continued in written form. In this way misunderstanding can be clarified and opinions fully expressed.

B. Function and Characteristics of a Test.

It is necessary to begin with a brief review of the psychiatric needs which give rise to testing. First of all, in order to help in diagnosis and in making decisions about treatment and disposal, the clinician requires information about the patient's past and present efficiency, about his attainments, personality and range of interests. Such information is as important for patients with clear-cut organic disorders as it is for patients with behaviour disorders. Even when such information is available, it is often difficult to obtain quickly, and often subject to the distorting effects of subjective factors on both the clinician's and other people's judgments.

In addition to the descriptive demands, the clinician finds it useful to have information of directly diagnostic import. For example, it may be useful to know that certain score patterns on motor tests are seldom to be found in neurotics, but are frequent in psychotics, or that certain types of results in the Block Design Test are only produced by brain-damaged patients.

Finally, the clinician often has to have independent indices of the course and effects of treatment. For example, it may be important to measure the changes in the level of intellectual ability of a paretic receiving penicillin treatment.

To fulfil such needs, the procedures used by the clinician must have three essential characteristics :

i. They must be objective, in the sense that they are as independent as possible of the clinician's changing moods and experience. Standard situations and standard methods of evaluating responses are necessary.

ii. The clinician must be able to compare a given performance of a patient with that of mentally healthy subjects, and of subjects having different degrees of the patient's own and other illnesses.

iii. Finally, it is necessary to know as much as possible about the factors affecting the performance of the task in question, and to what extent it really does involve the function under investigation.

We have here, in effect, three essential characteristics of a test : it must be objective, standardized, and validated before it can be brought into clinical use.

As soon, however, as we bring the test into clinical use, new problems arise, and it is these problems which will now be discussed.

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

C. Difficulties Arising in the Use of Tests.

1. *The Problem of Error.*

Error is inherent in all measurement, and it is necessary to know its extent. This would appear to be a statement of the obvious. A survey of some of the current literature dealing with diagnostic testing and discussion with clinical psychologists soon shows, however, that it is necessary to introduce the discussion of these problems into the field of clinical psychology. For example, many psychologists will confidently use the items of the Terman-Merrill Binet as indicative of abilities such as spatial imagery, memory, and so on without regard for the fact that the reliability of these items has been shown by McNemar (6) to be extremely low and that their validity has not been demonstrated. On the other hand, the discussion of error by the non-clinical psychometrician is often remote from the real problems of clinical work.

For the clinician the problem is a relatively simple one ; it is that a given result on a test may, for a number of reasons, give a seriously incorrect impression. Take unreliability as an example. Let us assume that we have a test with a test-retest correlation of ·9 and a standard deviation of 16. These data mean that, out of every three children obtaining an average score of 100, one would obtain a re-test score above 107 or below 93.* On the basis of the same data, out of every 10 children obtaining an average score on the first test, one child would obtain a score above 112 or below 88 on the second test. (These two illustrations assume that no learning has taken place.)

It is conceivable that, for the purpose of group selection, the relatively small error of misclassification arising out of a correlation of ·9 would be preferable to the much larger error resulting from unstandardized and unvalidated procedures. For the clinical worker, however, error takes on a different aspect. The psychologist does not wish to make a serious mistake with any patient, and he needs to know when he is dealing with that one person in ten.

The error of unreliability is not, of course, the only kind of error that arises for the psychologist. It is the error of misclassification that arises from imperfect validity which is crucial. It will not be necessary to labour the point that a test of perfect reliability but very low validity is not of much use to the tester. Validity coefficients as high as ·9 are very rare for tests in current use, so the chance of being seriously wrong about a patient on the basis of the test results is far too large for comfort.

The psychologist's worries are not ended, unfortunately, by a consideration of the errors of unreliability and invalidity. A third source of difficulty arises from the apparent impossibility of obtaining exactly representative samples of the population for the standardization of tests. This results in some distortion when considering the scores obtained on two different tests of the same function, or when considering scores in the same test given at different ages. The possible range of this kind of error is indicated by the data published by Dearborn and Rothney (3). They gave nine different group tests of intelligence and the Stanford-Binet to 320 children over a period of 12 years. The

* The formula for the standard error of estimate is used: $\sigma_e = \sqrt{1 - r^2}$. In the above example $\sigma = 16$ and $r = ·9$.

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

median score on the Binet was 102. The lowest median score obtained was 94 on the Otis Intermediate Form A, and the highest median score was 110 on the Haggerty Intelligence Examination, Delta 2. While the results of this study are by no means conclusive, they at least indicate the possible range of error that might arise from the differences in the samples on which the tests were originally standardized.

2. *Inadequacy of Information.*

A further complicating factor for the psychologist is that, not only must he take into account the possibility of considerable error, but more often than not he is in no position to do so because the makers of the test do not provide the necessary information. For example, the Terman-Merrill Binet is the only test on the market for which reliabilities for most levels of I.Q. and three main age-groups have been estimated. In contrast to this, many test constructors are content to publish one or two test-retest correlations. Some well-known British tests have been published without measures of dispersion, with the result that it is impossible to obtain a precise estimate of a subject's standing on the test. In any case, most of the tests in clinical use have been standardized in America, and have not been re-standardized for British populations.

The general effect is that the clinical psychologist not only has to deal with the problem of error, but is not given sufficient data on which to estimate its amount.

3. *Difficulty and Expense of Production.*

The psychologist's difficulties are increased by the fact that, not only are his instruments subject to error, but he very frequently lacks instruments which he badly needs. This arises from the fact that a well standardized and validated test takes a long time to make; and no matter how well a test has been standardized for the general population the psychologist who is serving a particular area in the end needs tests which have been standardized for that area. A psychologist doing clinical work rarely has the opportunity or the time to make the appropriate instruments for himself. Such efforts are costly and time-consuming.

4. *Inherent Limitations.*

There is a final difficulty which emerges sooner or later for the clinical psychologist. This arises from the fact that some questions cannot be answered at all by the procedure of the standardized validated test. For example, we may want to know how much of a patient's deafness is due to physical factors and how much due to hysteria. The fact that the patient obtained a maximum score on a hypothetical test of "hysteria" would not give an answer to this question. However, an *ad hoc* experiment carried out with the P.G.R. might give us a fair indication of the extent to which the actual deafness was being influenced by hysterical factors. The P.G.R. could be used to indicate differences in responsiveness to emotional, meaningful and meaningless stimuli, which could be presented from the range of tones which the patient claimed he could not hear.

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

Again, the question as to whether a patient's tic is increased or decreased by a condition of tension may help to answer a diagnostic question. Here, as before, the question is answered, not by a test, but by an experiment designed to give a specific answer. This would take the form of placing the patient in conditions of varied tension, and observing the changes in the frequency of the patient's tic in relation to these situations.

This kind of difficulty means that there are descriptive needs other than those described in Section B above. The clinician frequently needs to know the precise factors, and their inter-relation, which have produced a given symptom or disorder in an individual case, and the standardized validated test often cannot provide an answer to such questions.

In summary, it would appear that there are four main difficulties in the use of tests. There are first of all the difficulties of error. Secondly, we have the fact that for most tests the necessary information is not available which would allow us to estimate the amount of error. Thirdly, there is a shortage of necessary tests and the difficulty and expense of making them. Finally, there are inherent limitations to the use of the standardized validated test. Certain questions important for diagnosis cannot be answered by this technique.

These difficulties do not mean that we should discard the standardized validated test for clinical purposes. The psychiatric needs which give rise to the use of the test will remain, and so do the requirements for fulfilling them. The psychiatrist still needs tools for the effective description and diagnosis of his patient and an independent measure of the course and effect of treatment. Consideration of objectivity, validity and standardization must inevitably operate in answering such questions. What has happened here is that, as in many other fields, the solution of a certain problem results in the clearer emergence of others, and it is towards the solution of these other problems that we must now turn our minds.

D. Various Methods of Using Tests.

The reaction to problems arising out of the interpretation of test results in the clinical field has been varied, and it is now proposed to discuss some forms which it has taken.

1. *The "Qualitative" Approach.*

One development has been to differentiate between supposedly mere "quantitative" and "qualitative" aspects of mental testing. An example of this is provided by R. Schafer (8) on p. 17 of his book. He states: "A test response is not a score; scores, where applicable, are abstractions designed to facilitate intra-individual and inter-individual comparisons, and as such they are extremely useful in clinical testing. However, to reason—or do research—only in terms of scores and score-patterns is to do violence to the nature of the raw material. The scores do not communicate the responses in full. For example, in response to the question 'Why should we keep away from bad company?' both of the following responses obtain a maximum score of 2: 'They have a terrible influence on our character and lead us into evil ways,' and 'I don't have much use for the concept bad company but we're supposed

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

to believe that they will corrupt us.' . . . The subject communicates more than he wittingly intends; he also communicates more than can be scored. Test responses, because they represent the subject's style of thinking, allow for inferences concerning predominant features of character make-up."

The counterposing of the qualitative versus the quantitative approach in mental testing appears to be based on two fallacies. The first fallacy is that, because an observation is "qualitative," it ceases to be subject to the rules of reliability, standardization and validation. In fact, to observe an hysterical quality of a patient's response does not necessarily mean that the patient has an hysterical personality, unless evidence has been obtained which validates this notion. Nor does it mean that different observers would necessarily judge the response to be equally hysterical; or that, if they agreed in their judgment, the patient's response was any more hysterical than that of any other person.

A second misconception implied by this approach is that qualitative observations cannot be quantified. In fact, the quantification of qualitative observations in an objective manner is primarily a question of ingenuity and the level of technical development of science. For example, it would be possible to make up a number of questions, of the kind mentioned above by Schafer, which would provide an opportunity for measuring hysteric-like responses. A definition of hysterical quality could be set up as a scoring criterion and the test could be administered, resulting in a quantitative score in terms of the number of hysterical responses. These would then be subject to the usual psychometric treatment.

It seems, therefore, that when a psychologist generalizes from hysterical features in a patient's responses to the Comprehension subtest, he is in fact using an unstandardized and unvalidated test. The only use for such observations, therefore, must be that they serve as hypotheses to be tested by proper investigation.

2. The "Flexible" Approach.

Many psychologists who look upon the testing situation as a method of evoking qualitative responses have gone much further than Schafer does. It was quite clear at the 1950 General Meeting of the British Psychological Society, for instance, that some psychologists do not believe it necessary to give a test in exactly the manner demanded by the designer of the test. They claimed that this imposes a restriction on the investigation of a patient, and that the test should be used in what might be called a "flexible manner." This approach is subject to a number of criticisms and qualifications. In the first place, the claims to get better results by so-called more flexibly used tests is subject to the same needs of validation as any other claim. Precise evidence would have to be supplied that such methods produce as good or better results than the more rigorous methods.

A second criticism is that this approach implies that the psychologist must choose between the rigorous standard approach to the patient, and the flexible one. In fact the psychologist finds it necessary to use both methods. The main question is to decide on the appropriateness of the given method.

If the question concerns the patient's relative standing on a test, then any

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

change of procedure must be avoided. If, on the other hand, the psychologist is searching for hypotheses to account for some of the patient's responses, he might find it useful to conduct an uncontrolled inquiry. For example, the psychologist may believe that the patient's responses in an individual test are being affected by his nervousness. This suspicion should not affect the actual content of the situation presented or the scoring of the responses. When the test is over and the patient appears to be more settled the psychologist might well return to some of the items involved, and ask the patient some questions about them to see how much and what kind of help is necessary to obtain a correct answer. Such inquiry would provide the basis for further investigation.

It is usually only in extreme cases, where no responses at all have been obtained, that procedures have to be modified. Such modifications should be minimal, just enough to make the difference between no response and some response. The usefulness of the free inquiry does not justify the claim that loosely controlled inquiries are valid. Without any proof, this can open the door to dangerous practices—dangerous, that is, to the interests of the patient.

The third weakness in this approach is the assumption that a completely free approach is the only way in which the full co-operation of the patient can be obtained. In fact, one of the main skills of the psychologist lies in his ability to obtain the co-operation of the most intractable patient under conditions which depart as little as possible from those laid down.

The last two points are illustrated by the following examples: The psychologist had to give a Binet to a 4-year girl whose main complaint was that she did not talk. He hit upon the expedient of placing the little girl on his knee and giving her a kiss for every item which she completed, regardless of whether it was right or wrong. Every item was presented and every response scored rigorously according to the instructions in the manual—with the one exception of the reward for responses. The patient completed enough of the test to show that mental defect was not to be considered in her case, and she walked out of the testing room talking to and kissing everybody that she met. A second example was an adolescent girl who would not talk and was thought to be a mental defective. The patient was persuaded to indicate by a slight movement of the finger which of the items in the matrices was her choice. Here again she chose enough of them correctly to indicate that she was not a mental defective, but that she was essentially a psychiatric problem. In both of these cases the instructions and the scoring methods were adhered to as closely as possible, and the findings were of undoubted value in the diagnosis of the patients' disorders. If the adherence had not been so close, the contra-indications of mental defect in these cases could not have been so definitely made.

3. *Experimental Method.*

It would seem that the correct reaction to the difficulties inherent in testing would lie in the psychologist's trying to find a way in which he can carry out his descriptive and diagnostic functions more effectively. It is the main thesis of this article that this can best be done by making explicit the underlying principles of these functions, by formulating in general terms what every clinician does when he is investigating a patient.

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

First of all, the features presented by the patient, his life history and medical history and test results are integrated. In this the discussion between the psychiatrist, social worker and psychologist play an important part. As a result certain formulated problems emerge.

The next move is to advance hypotheses which will explain them. This is in essence what the clinician always does, though at times he will not be inclined to admit it.

The third step the clinician takes is to ask himself what effect the truth of each of his hypotheses will have on the treatment and disposal of the patient. If any of these hypotheses are not likely to have an effect on treatment and disposal, he will be disinclined to test them.

Finally, the clinician has to decide how he is going to test the various hypotheses.

We would submit that these four phases are, in fact, always carried out when a clinician makes a diagnostic decision. Making them explicit helps us to evaluate the various procedures which are used in diagnostic work, and it helps to carry it out more effectively and economically in terms of time. When we turn again to the standardized validated test it becomes immediately apparent that it is not the only way of testing a hypothesis. The ways of doing this are as varied as the whole range of human observation. For example, one hypothesis to explain a pre-school child's disorder might be that it is a reaction to the restrictiveness of an over-tidy, meticulous, houseproud mother. If, however, one visits the home and finds that it is untidy, neglected-looking, and that the mother is on her way out to go dancing with a man who is not her husband, then the hypothesis is strongly contra-indicated. Similarly, the hypothesis that a child is over-inhibited may be contra-indicated by mere observation of the child in his relations with other children. On the other hand, if a validated and standardized test of inhibition were available, the answer to the same question might well be available more quickly and more unequivocably than, say, the observation of the child's behaviour. The methods open to the psychologist are many; the essential characteristics of the investigation are, however, always the same.

E. An Example of the Application of the Experimental Method.

In the interests of brevity, the case analysed here is to some extent a hypothetical one. The patient is a 10-year-old girl who comes to the hospital suffering from enuresis. Other information about the patient presents a picture of a quiet, rather retiring girl who has done averagely well at school and has shown nc outward behaviour problems. In the course of the admission routine she was given the Terman-Merrill Binet; she obtained a quotient of 120 and a score of 7 words correct on the Vocabulary. Statistical computations show that this discrepancy is rarely to be found in the normal population.*

* The formula used was (as provided by A. Lubin):
S.D. of difference between I.Q. and Vocab. $= \sqrt{\sigma^2_B + \sigma^2_V - 2r\sigma_B\sigma_V}$,
where σ_B = standard deviation of the Binet,
σ_V = ,, ,, ,, Vocabulary,
r = correlation between Binet and Vocabulary.

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

The data can now be classified in the following way: The child is superior in her Binet performance, average or unexceptionable in her school work and general behaviour, and exceptionally bad in the Binet Vocabulary. The main problem is formulated in this way because, as psychologists, we have to look at the problems mainly from the point of view of our test results. Four hypotheses are put forward:

(1) The patient is an extremely intelligent child who comes from an abnormally poor social and economic environment, resulting in the poverty of her vocabulary and in frustration which may be a factor in her present disorder.

(2) The patient is an extremely intelligent child who suffers, for an unknown reason, from an extreme degree of emotional inhibition which results in breakdown in a controlled interview situation like the Binet Vocabulary test. However, her intelligence is great enough for her to be able to get by in most situations to which she is accustomed.

(3) The low Vocabulary score is due to a temporary, but extreme disturbance coincident with her arrival at the clinic, if not related to it, and would never be repeated. It therefore has no diagnostic significance. The Vocabulary test was the first given to her and was, in fact, the first effective contact with the clinic.

(4) The inability to express herself verbally is due to some kind of specific aphasic defect in a child who is of bright, but not necessarily as great intelligence as would be indicated by hypotheses (1) and (2). This defect may or may not be related to the patient's disorder.

Note that none of these hypotheses are mutually exclusive, nor do they necessarily account for the disorder.

The next step in our procedure is to examine each of the four hypotheses to see what effect the truth of each would have on the treatment and disposal of the patient. It is obvious that hypotheses (1), (2) and (4) would considerably affect our lines of treatment. Hypothesis (3), if it were true, would result in our closing the investigation of the test results, and no further contribution to the case could be expected from this particular point. It is worth our while, therefore, to investigate all these hypotheses and so we move to the next stage, that of designing tests for each one of them. The first hypotheses can be tested by direct inquiry from the social worker on the basis of a visit she has paid to the home. If she finds that the home is a superior working-class one, with varied cultural interests, and the child herself spends a large proportion of her spare time reading books like the *Children's Encyclopaedia* and children's classics, then we can fairly say that the hypothesis has been disproved.

Hypothesis (2) can be taken up in a variety of ways, depending on the amount of time at our disposal. If we are extremely busy we may look for a merely confirmatory test, such as giving the child the Mill Hill Vocabulary Test, together with a repeat of the Binet Vocabulary. The Mill Hill Vocabulary is a group test and the child can be left alone to answer it, thus eliminating the personal face-to-face situation which is supposed to have lowered the Binet Vocabulary performance. If the score on the Mill Hill Vocabulary were significantly higher statistically than that predicted from the Binet Vocabulary, the hypothesis would be confirmed but not proved. There may

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

be other reasons why the child finds the Mill Hill type of test easier to do. With more time at one's disposal, one might set up an experiment in which the individual and group type tests are given in some kind of balanced order. Here again, if the results are in the direction predicted, they would only be confirmatory. A third way of handling the situation would be to use a valid test of inhibition. Such a test, while its finding would be of direct clinical use, would still only be of a confirmatory character because it would not necessarily prove that there was a causal relationship between the child's inhibition and performance on the Binet Vocabulary. The only way in which we could make a crucial test of the hypothesis would be if we were able to set up the inhibition as an independent variable and vary its intensity in the patient, at the same time measuring the effect of such variation on the dependent variables: individual and group type vocabulary tests.

It must be remembered that the results arising out of the first testing of an hypothesis may result in the reformulation of the problem and the devising of a new hypothesis. An investigation becomes, therefore, a process in which every new fact results in the re-formulation of hypotheses and every re-formulation can result in the unearthing of new and relevant facts. Thus the work of the psychologist does not stop at the first formulation of problems and hypotheses.

F. How the Experimental Method helps to Meet the Main Difficulties.

It is now necessary to review the earlier statement of the difficulties arising out of the use of the validated standardized test to see to what extent the conception of experimental method really meets these difficulties. How does it make up for the inadequate information about the tests and the lack of suitable tests? Finally, does it enable us to answer questions about the precise relationships of factors within an individual patient?

1. *The Problem of Error.*

All methods of scientific investigation are, of course, subject to error. Crucial tests for hypotheses are provided with relative rareness: so, while the results may be consistent with an hypothesis, the underlying causes may, in fact, be different from those suggested. Any conclusions, therefore, which are drawn from such data would be in danger of error.

The experimental method, however, does help to reduce a good deal of the error arising from the use of ordinary standardized validated tests. If, for instance, in the detailed example described above, the first test results were to be taken strictly at their face value, they would merely indicate that the child was of superior general intelligence but that she had an extremely poor vocabulary. Actual investigations did show in this particular case that her vocabulary was, in fact, commensurate with her Binet test score, but that it was the combination of her own personality and the peculiarities of the test situation which depressed her test score considerably. It is the investigation into the observed discrepancy which, in fact, prevents the drawing of incorrect

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

conclusions about the patient; and thus we can say that in this case the use of experimental method reduces the extent of error which arises from the naïve use of standardized validated tests.

2. *Inadequate Information and Difficulty of Standardization and Validation of Tests.*

The inadequate information about available tests and the difficulties in making tests which are needed were the second and third difficulties mentioned in Section C. Here again the use of experimental procedure helps us to overcome such difficulties. This is illustrated by a patient who was a deaf mute, and whose difficulties were not at a sensory, but at a perceptual level. One question which arose was that of his trainability. No test of "trainability" for this kind of patient exists. This question was answered by training him to carry out various tasks under controlled conditions. Where before he had scored zero on Raven's Matrices and the Porteus Maze, he now performed such tests at about average level. These results seemed to contraindicate the diagnosis of schizophrenia or general mental defect, and gave positive indications of his "trainability."

3. *Overcoming the Inherent Limitations.*

It was pointed out in Section C that there were inherent limitations in the standardized test procedure. Examples were given of a patient with hysterical deafness and a patient with a tic to show how, in fact, a hypothesized causal connection between the symptom and the variable can be tested by an experiment. Such a connection, of course, cannot be demonstrated by the ordinary test. The case outlined in Section E above provides another example of the way in which answering the question posed by an individual case requires the testing of specific hypotheses about it. A standardized validated test of inhibition, while it could give results which have a bearing on the main hypothesis, would not show to what extent the child's inhibition was in fact related to her low Vocabulary score and in turn to the enuresis. It is direct experiment which might give a much more accurate indication of this.

G. Outcome of the Experimental Method.

1. *Testing only to Answer Definite Questions.*

The general approach described above means that no test is given to a patient—in fact no inquiry of any kind is made—unless a definite question is being answered, i.e., a definite hypothesis is being tested.

This approach is, of course, in contrast to the battery approach favoured by many psychologists, and which appears to be characteristic of the work of some American clinical psychologists. For example, Schafer, in considering the main propositions in the clinical application of psychological tests, states: "The subject must be made to think in a variety of problem situations to enable the examiner to distinguish the pervasive, fundamental or pathological aspects of his characteristic adjustment-efforts. In any one situation he may

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

present a one-sided picture ; in a variety of situations—in a *battery* of tests—a rounded and hierarchical picture can be expected " (8).

Actually, completion of a " rounded and hierarchical picture " does not necessarily provide any check against error. The battery of tests is just as much subject to the demands of validation as a single test, and each additional test must be shown to contribute to predictive efficiency.

There can only be three reasons for giving a patient a test : either it has been shown to differentiate certain qualities of validated importance for diagnosis, treatment or disposal, with a known degree of error ; or there is agreed and long clinical experience indicating the possibility of the validity of the test ; or, finally, the test is deduced from certain specific hypotheses entertained about the patient. The notion of validity is inherent in all testing work.

It does not follow from this, however, that the routine test battery approach has no place in clinical work. Procedures must be routinized when it is found that certain questions nearly always arise in the course of the investigation of a patient, and that validated procedures exist which can give the answer to these questions. For example, under present conditions, the important question that always arises with children is the level of their various abilities, and therefore it becomes necessary to make intelligence tests part of initial clinical routine. If gradually the number of such questions which can be answered in this way increases, then we will have the development of a battery of tests.

However, there seems to be little advantage in developing a time-consuming battery approach utilizing unvalidated tests to verify hypotheses which no one has advanced. It would seem to be far better to spend the time validating by experiment specific hypotheses about patients, or to carry out long-term research which would enable us to inaugurate useful routine procedures.

2. Limiting " Test Interpretation."

Another approach which is, of course, related to the battery approach, is that which rests mainly on the integration of all available data about the patient. Rosenzweig (7) has explicitly formulated a rationale and a method of such an approach : " Taking it for granted that all expressions of personality are, if properly understood, meaningfully related to each other, the interpreter attempts to single out the cardinal strengths and weaknesses, drives and interests, attitudes and symptoms in terms of which the various manifestations of the person seem to be most significantly oriented."

The method which he uses in operating this approach is the method of " minimum hypothesis," which he described in the following way : " The picture of the subject that will in the end be embodied on the psychological report must account for the largest number of underlying or explanatory variables. When two alternative interpretations are in competition, the foregoing statement provides the only currently accepted basis for deciding between them."

This use of the " minimum hypothesis " would seem to be an application of the law of parsimony. This can only apply when all the available evidence has been gathered and more than one hypothesis can satisfactorily account

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

for *all* of the evidence. If, however, the principle is used merely to analyse the evidence which has been gathered up to date, and becomes in effect a reason for relinquishing further investigation, then the principle of the "minimum hypothesis" becomes a factor which limits the effectiveness of the psychologist's investigations. The best rule for the psychologist would seem to be that, if he has more than one hypothesis which can adequately explain the data which he has gathered, he should perform a crucial experiment which would enable him to decide which of the hypotheses would, in fact, account for the data. Merely to present an "explanation" to the psychiatrist is not good enough. An unsubstantiated explanation is an untested hypothesis.

This does not mean, however, that the psychologist must not carefully relate all his available data pertaining to the patient's condition and life-history on the one hand, and the available information about the construction and validation of the test he has used on the other. Nor does it mean that the psychologist may not frequently find himself unable to test rival hypotheses; even in such circumstances he might not, in fact, use the principle of parsimony. His choice of hypothesis may not depend so much on which hypothesis has the smallest number of variables; but the hypothesis which, while explaining the largest number of facts, also enables the psychiatrist to make a disposal which is administratively convenient; or which, say, best suits the patient's personal situation. The principle of parsimony, therefore, is a relative one, and the way it is used depends on the purpose of the psychologist. It is not a principle which must necessarily yield the truth in a situation where few of the relevant facts have been established.

3. *The Abandonment of the "Artistic" Approach.*

Many psychologists insist that the clinical method is something different from the ordinary methods of science. For example, T. G. Andrews (1) states: "the clinical methods have a different frame of reference from the experimental and differential methods of psychology." For further definition of this third method he refers the reader to the article in this book by the late Andrew Brown (2).

While in most respects the description of the work of the clinical psychologist in this article is apposite, to talk as it does of the "art" of clinical work as one of its essential characteristics is to lay oneself open to many dangers. Such an approach is, in effect, a way out for those who will not or cannot make their hypotheses explicit and who avoid seeking rigorous tests of such hypotheses. It is, of course, legitimate to decide to work on the basis of a certain assumption with a patient because the patient's condition urgently demands action and the psychologist or psychiatrist knowingly takes a step into the dark. There are relative degrees of certainty with which one can entertain theories about a patient's condition and what can be done about it. However, it should always be possible to formulate in advance the possible outcome of one's theories, so that the actual can be compared with the predicted outcome. The fact that the applied scientist has often to plunge into guesswork and trial and error does not warrant the description of his work as being essentially an art, or a matter of intuition. To insist on something more than

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

science, rather than admitting plain but explicit guesswork, or to make a virtue of the fact that as yet we do not understand the processes of creative thinking, is to guarantee confusion, even though brilliant discoveries might be made in the course of such work. The onus still remains with the psychologist to use his ingenuity to make his thinking and procedures explicit, so that he formulates tests of his hypotheses and increases the efficiency of his work and enables others to learn from his failures and successes.

4. *Qualifications of the Goldstein-Scheerer Approach.*

Goldstein and Scheerer (5) state that "a patient might produce a minus result on a subtest without the function presumably tested by the applied subtest necessarily being lacking, and he might produce a plus result without the function presumably being tested necessarily being present." They state that this happens very frequently, and they give two examples which are worth examining. The first is that of a patient with an apparent alexia. When, however, a frame is placed around a word, that word is correctly perceived. Another example is that of a patient who was unable to do a sentence-completion test on one day but did a different sentence-completion test perfectly on a subsequent day. The explanation advanced for this finding is that on the second occasion the sentence which had to be completed had a personal bearing on the patient's present situation, whereas in the first test the content had no bearing whatever. They conclude, therefore, that the first failure was not due to a real inability to do sentence completion, and that therefore a zero score would have been misleading.

This point of view, and the examples given, are in agreement with one of the main theses of this article, i.e., that the test score does not necessarily indicate the patient's true standing on the function which it is supposed to measure. This fact does not justify Goldstein and Scheerer's conclusion that quantitative scoring and qualitative analysis are methodologically opposed.

First of all, ordinary quantitative methods are inherent in the examples quoted. The patient in fact scores zero in one kind of situation and full marks in another, and the discrepancy is probably statistically significant.

Once such a discrepancy has been observed an explanation has to be found. Goldstein and Scheerer put forward their own explanations, but they are not necessarily the only explanations that could be put forward, and not necessarily correct. These explanations would have to be validated by experiment.

In the light of these considerations one would therefore not pose the qualitative against the quantitative approach to diagnostic work. Quantification is usually necessary for exact observation, while qualitative analysis is always necessary in the development of hypotheses to account for one's observations. In any case, the demand for validation can never be avoided.

H. Obstacles in the Way of Developing New Procedures.

It is one thing to point out the snags in a well-established procedure, and the general lines along which they might be overcome. It is another thing to carry out such a line in practice. In fact, in the field of diagnosis the method of the experimental investigation is almost non-existent. The application

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

of test methods for measuring the course and effects of treatment and in the disposal of patients has hardly yet begun. Among the obstacles to an effective development of the quality of the work, the most important would seem to be pressure from immediate colleagues, general shortage of time, and, perhaps, psychologists' own attitudes of mind.

1. *Pressure from Colleagues.*

There is considerable pressure from psychiatrists and social workers for quick and understandable results which serve as a short cut in the solution of difficult problems. There is a standing temptation for the psychologist to answer a query with " Yes, certainly. I'll give him the such-and-such and the so-and-so tests." His colleagues need not know that the standardization and the validation of these tests are practically non-existent ; and that, in fact, in this particular case no test can be a substitute for a careful investigation of the patient's school and work history, which should be evaluated by someone trained in educational and industrial psychology.

2. *Understaffing.*

A second factor is the shortage of time. It is sometimes necessary to spend 30 hours on a patient before the main questions can be answered. The premium on time is due mainly to the large case loads and the relative understaffing of psychological departments. Until psychological departments are increased, it is necessary to review some of the more time-consuming traditions, such as that of the Binet for children and the apparently unqualified virtues of " individual " testing for all patients. Psychologists and psychiatrists need to review very critically the way they conduct their work, and to what extent the time spent on patients answers real questions. It may well be that the best approach consists of putting patients through a few strategically chosen group tests carried out by non-psychologists, so that the psychologist can devote time to intensive work on individual patients, or on fundamental research.

3. *Misconceptions.*

A third obstacle has been that of the attitude of mind of many clinical psychologists. One attitude implies that there is no scientific method of dealing with diagnostic problems other than that of the standardized validated test.

Psychologists who share this belief fall into two main categories. First, there are those who appear to believe that the only descriptive and diagnostic service that the psychologist has to offer is the provision of a test result. Secondly, there are those who are impelled by the limitations of routine testing to justify less scientific methods of approach. It is obvious that neither category is likely to play a vigorous part in the development of diagnostic experiments.

A second factor which hinders many clinical psychologists is a misunderstanding of the effect of multiple causation. They believe it is impossible, for

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

example, to find out by experiment why a patient cannot read, because of the multiplicity of factors which can produce a reading disability in a given case. The " causes " one looks for, however, depend on one's purpose. With a non-reader, for example, the main question is what is preventing him from reading *now*. One likely hypothesis is that he now has no physical or cognitive handicaps. This hypothesis can be tested by actually teaching him to read. If he succeeds in learning to read, then it would not be sensible to be paralysed by the fact that one cannot unravel the network of causes which prevented the patient from learning to read in the past.

This does not mean that the findings of large-scale investigations should be neglected, since they provide the basis for formulating hypotheses about a given patient and for developing long-term plans for preventive measures. The main point for the clinician is that the data provided do not necessarily give an answer to the question of why a given patient cannot read, nor do they exclude the possibility of finding an answer to such a question.

The main consequence of these difficulties appears to be a relative lack of integration between general psychology and clinical psychology. The clinical psychologist does not find himself forced to apply what he knows of the processes of perception, learning, thinking and feeling to his work. There is no intense process of mutual enrichment between general and clinical psychology. It is only occasionally, as in the work of Bender and Teuber, Gelb and Goldstein, that a two-way relationship between general and clinical psychology is shown.

I. Discussion.

It must not be assumed that the experimental approach has not featured in both the writings and the work of many clinical psychologists. Most of the prominent psychologists working in this field have in one way or another described this approach, and every psychiatrist and neurologist uses this method in some way or other.

The Goldstein-Sheerer Monograph (5) gives examples of control experiments—for example, in the Colour-Form and Object-Sorting tests, which are implicitly based on the same principles as those described in this article. In his book Goldstein (4) suggests that the examination of patients should be taken to the point where " we can predict with relatively great certainty how he (the patient) will behave in any situation, even in respect to tasks which we have not yet examined. Only cases which have been investigated with such thoroughness should be used in the formation of a theory." A rather similar point is made by Schafer (8) when he says: " an *interpretation* is a prediction that certain phenomena of behaviour or thinking will be found by direct observation to characterize the subject. . . . An interpretation should be subject to immediate validation by refined observation. One implication of this definition must be stressed: the prediction must deal with a *characteristic*, with a mode of behaviour or thinking that is enduring and that sets the subject off from many or most other subjects."

Rosenzweig (7), too, states that there are conditions when he would formulate and test hypotheses about a patient. This is when the " insightful com-

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press

prehension demanded of the clinician cannot be achieved." He concludes, " every individual becomes a research project in his own right."

In Zubin's (9) article in T. G. Andrews' *Methods of Psychology* there is given an example of the experimental investigation carried out on an hysteric, though the purpose of this experiment was not a diagnostic one.

In general, one can say that there is nothing new in the idea of the use of experimental method for diagnostic purposes. All that can be said to be new in this presentation is that experimental method is made the main approach to all diagnostic problems, and the standardized validated test thus takes its place as one very important method of testing hypotheses.

J. Concluding Summary.

This article was concerned with the discussion of only one aspect of the work of the clinical psychologist : the use of psychological tests for diagnostic purposes.

It discussed the needs which give rise to the use of the standardized validated test, and the difficulties that arise when the results are being evaluated.

The main thesis put forward was that the difficulties can, to a large degree, be met by making explicit the principles underlying all clinical investigations and that these principles are those of experimental method. Psychological tests, therefore, can be seen as one method, but not the only one, of testing hypotheses about a patient.

My thanks are due to Mr. A. Lubin and Dr. J. G. Ingham for their interest and criticisms.

References.

(1) Andrews, T. G., *Methods of Psychology*, 1948. London : Chapman & Hall, Ltd.
(2) Brown, Andrew W., " Methods and Techniques in Clinical Psychology," *in* Andrews, T. G., *Methods of Psychology*.
(3) Dearborn, Walter F., and Rothney, John W., *Predicting the Child's Development*, 1941. Cambridge, Mass. : Sci. Art Publishers.
(4) Goldstein, Kurt, *The Organism*, 1939. New York : Amer. Book Co.
(5) *Idem* and Scheerer, Martin, " Abstract and Concrete Behaviour," *Psychol. Monograph*, **53**, No. 2.
(6) McNemar, Q., *The Revision of the Stanford-Binet Scale*, 1942. Boston : Houghton Mifflin Co.
(7) Rosenzweig, S., *Psychodiagnosis*, 1949. New York : Grune & Stratton.
(8) Schafer, R., *The Clinical Application of Psychological Tests*. Menninger Foundation Monographs, No. 2.
(9) Zubin, J., " Objective Studies of Disordered Persons," *in* Andrews, T. G., *Methods of Psychology*.

https://doi.org/10.1192/bjp.97.409.748 Published online by Cambridge University Press