

Chapter 2

Development of Psychological Testing in India

S.K. Verma

1 Introduction

When I joined the Department of Psychiatry at the Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh, many of the psychological tests used in most parts of the country suffered from several deficiencies, known and unknown till then. In fact, one cannot work in an area where the tools in use are deficient in many ways. If one has to work honestly in this area, one cannot hide behind these imperfect tools, without doing anything about it. Blaming the tools alone cannot help one for long, particularly, when this happens to be one's chosen field of work and worship. The God is everywhere, seeing everything and forgiving nothing.

But, before I name these deficiencies in tests in use then, I may be allowed to define an ideal test and mention its necessary characteristics. In all fairness, I must accept that no single psychological test can be considered as a perfect test for all types of populations, for all subjects and for all times to come. But, realising our own limitations in knowledge, wisdom and availability of time and resources, a psychological test has to have most of these desirable characteristics, while leaving scope for improvement in the future.

S.K. Verma, Former Additional Professor of Clinical Psychology

S.K. Verma (✉)

Clinical Psychology, Department of Psychiatry, Postgraduate Institute
of Medical Education and Research, Chandigarh, India
e-mail: savita.pgi@gmail.com

2 Definitions

A psychological test is a standardised method of (a) comparing the behaviour of two or more persons at the same time and/or (b) comparing the behaviour of the same person in different periods of time.

A standard is that in terms of which we measure something, e.g. a gram, a litre, a mile, a rupee or an inch; measurement is attributing a number according to well-defined rules.

Behaviour includes both overt and covert behaviour, observed and expressed verbally and/or measured.

Standardisation is a process of exactly fixing the stimulus variables, the exact situation of administration and of observing the behaviour, in terms of well-defined units of measurement and the exact method of interpretation, in the light of relevant norms, having high reliability and validity.

Objectivity refers to freedom from subjective factors in administration, measurement and interpretation.

Reliability means consistency of scores over a period of time and consistency in administrative methods, administration by self or by others, the method used by scorers as well as the test's internal consistency. It makes one confident that every time a test is used, the same test is used and similar methods are used for administration, scoring and interpretation. Reliability can be of the total score, of various parts of the test or of every individual item. It is mostly measured in terms of coefficients of correlation, but can also be estimated in terms of percentage of agreement or disagreement.

Validity of a test shows whether a test measures what it purports to measure, e.g. a test may measure more than one aspects of behaviour, but one of them may be more specific.

Intelligence tests, for example, must measure intelligence in particular, but may also show the approach in solving a problem, a method of solving problems and how a person reacts to frustrations when difficulties arise while solving a problem. A personality test may also measure one or more aspects of personality, i.e. interest, attitude, preferences and aptitudes. A test is a valid measure of behaviour and may be valid for one aspect of behaviour only. The validities are also of different kinds: face validity, content validity, concurrent validity, construct validity, factorial validity and predictive validity.

Other features of validity of a test are as follows:

- A test is valid for something with which it correlates.
- Validity depends upon the reliability of the test. In fact, validity cannot exceed the under root of its reliability.
- A test may have different validities for different aspects of behaviours.

Norms have to be relevant and meaningful. Local norms have to be used for interpretation. Population, reliabilities and validities are defined by the standardisation process and must be given in the test manuals. Norms have to be specific with regard to location, age, sex, education and all other significant variables likely to influence test scores, reported or not reported so far.

Test manuals should ideally provide all the above information. No test can be culture free and culturally fair unless it is proved, reported and defined. Manuals can also provide cross-validation data, if and when available. That is why a test can be revised and has to be revised after sometime, in order to remain a valid and reliable measure over time.

3 Drawbacks or Deficiencies of Psychological Tests

The following drawbacks were most prominent when I joined the department in 1968:

- Most tests used were in English language only, and on-the-spot translations were being done on the bedside of patients in the general hospital setting. Uniform Hindi translations were needed, with local norms for use with them.
- Patients were illiterate or barely literate and unsophisticated in the use of tests. Some of them were even so frustrated that they were ready to give their thumb impressions, in place of responding to test items.
- Foreign norms were being used while scoring and interpreting the test scores, whereas there was the need to prepare local and relevant norms for use with our patient population.
- The tests used were too time consuming in terms of length of items, e.g. nearly 200 in the General Medical Index Health Questionnaire or the CMI Health Questionnaire (Erdmann Jr. et al. 1952) and 60 in Standard Progressive Matrices (Raven et al. 1998). The language used in Hindi versions of personality tests was complicated, with difficult words, e.g. “*ashrupat*” for tears in the eyes, as in the Hindi Personality Trait Inventory (Verma et al. 1990).
- Tests in simple Hindi were not available, which were constructed using patients’ own symptoms with their own expressions, and for which, local norms were available or which could be used at all educational levels.
- A need for modifications of some tests was felt, but no constructive and useful work was done, and at best, only lip service was paid. The problems were noted, but nothing was done to address them. There was a scope for construction of tests in many areas, but much needed work was missing in the absence of a team of clinical psychologists devoted to this work wholeheartedly.
- In a few areas, e.g. Bhatia’s battery of performance tests of intelligence in India (Bhatia 1955), where useful, constructive work was done with separate scoring systems for literate and illiterate subjects, there was lack of follow-up work with revision of the scoring system and revision of norms from different parts of the country.
- In addition to the above drawbacks, the attitude of the so-called experts from Indian universities was deplorable, to say the least, because test construction was looked down upon even for a PhD work. In my personal experience, while doing my PhD, the “experts” made one excuse after another for it—such as “It is not sufficient work for PhD” to even saying “One PhD is not sufficient for test construction, or “Test construction in India is not accepted for PhD,” or “Only in

North India some universities have accepted a few PhDs in test construction,” or “There is nothing original in test construction.” Even when evidence to the contrary was available, such excuses continued to be made.

I was lucky enough to have Prof. Wig (who accepted my arguments and allowed me to continue my PhD work) who accepted to be my supervisor and encouraged me to continue my experiments. (That is why I call him my godfather.)

- These “experts”, when I approached them for guidance or advice, were of the opinion that we should adopt standardised tests from the West and use them with our population to solve any research problem, a piece of advice, which I could not and did not follow, for obvious reasons. I was too junior a researcher, so I bowed my head in silence, but continued my work, the only way open to me at that time. I do not know what kept me going in those early years, despite the continued opposition by my seniors, colleagues and juniors, both psychiatrists and clinical psychologists, in this Department of Psychiatry (not of Clinical Psychology). But looking back over the years, I can name a few possible “reasons” for it.
- I saw construction of psychological test as writing beautiful poems as a hobby, enjoying the process, as my way of worship to the “God of Measurement”, if there is any. (If others can worship other Gods or Goddesses such as *Brahma*, *Vishnu*, *Mahesh*, *Lakshmi* or *Saraswati*, why cannot I have my own God and worship Him.)
- I had read somewhere the definition of morale (given by a veteran soldier) as “Morale is what keeps your feet going while your head says it can’t”.
- So I bowed my head and like a good student, started doing my work, as a worship in order to please him (i.e. the God of Measurement), to shower His invisible blessings (no matter if it is seen or appreciated by others or not). To be able to do one’s preferred work, is its own reward and can never be replaced by any other award or reward.
- Appeal of my heart—“Do not give up before doing your best, no matter what the consequences”. So I did continue my work as worship, on a war footing, if I can call it that.
- My father’s advice—“Whenever in difficult circumstances, bow your head in silent prayer and start doing what you think is right, without caring what others say or do”. My late father often referred to his own example—when he started his own work as a dentist, people made fun of a dentist as a mere tooth puller, as one who is destined to end up on the road calling people for repairing their broken teeth. But, he never gave up and ended as one of the leading dentists in the city of Varanasi at the time.
- Never-ending faith of my late wife in my ability to overcome my shortcomings, my difficulties and be successful in the end. This moral support proved to be of great help, whenever I happened to be in low spirits and seeing faces of my own children, as if asking me “Water water everywhere nor any drop to drink” (Coleridge in his “Rime of the Ancient Mariner”), as to what messages would my life and efforts leave behind, for them to follow. (“Keep your head above the sea level, in order to survive and win”.)

- It is rightly said that “All is well that ends well”. At the start of my professional career, opposition to my views was the rule; this was gradually converted into active cooperation by all concerned, from all parts of the country. My prayers were being answered after all and in my life time.
- When there were other PhDs, not only by clinical psychologists but also by psychiatrists of the department, it was more than I could ever expect or imagine or live to see.

In this regard, I must gratefully acknowledge the significant contribution of other fellow clinical psychologists, particularly of Dr. Dwarka Pershad, Ex. Additional Professor of Clinical Psychology, Department of Psychiatry. Indirect encouragement to our efforts was soon available in the form of a few national awards for such tests from professional bodies, including one by the Indian Psychiatric Society, and my own nomination to advisory bodies and editorial boards of various journals in the country. I admit here once again the full and constant support of Professor Wig, without which I would have left this great department long ago. Last but not the least was indirect factors such as constructive criticism by eminent professionals, foremost among them were Prof. N.N. Sen, Prof. G.G. Prabhu and Prof. J.S. Neki, and international authorities on tests such as Prof. H.J. Eysenck, Prof. R.B. Cattell and Prof. Guilford.

4 Our Contributions

Our contribution in this area has to be seen as an insignificant and tiny effort, in the universe of a vast area of research and possibilities. Doing something (however small it might be) is better than doing nothing in the area of construction of psychological tests. I had read somewhere (in a different context though) that:

A piece of bread is better than nothing.
Nothing is better than God.
So, a piece of bread is better than God.

If I may be forgiven, I would like to replace “a piece of bread” by a psychological test in this statement.

Similarly, in another context I have read:

O God, if there is a God.
Save my soul, if there is a soul.

If I may be forgiven, I would like to put it as:

O God of Measurement, if there is a God of Measurement. Accept our small offerings (tests as flowers); if they can be offered and be accepted.

We are not sure (at least I am not sure) that what we have attempted to do in the field of psychological tests would make a difference to the world of measurement and be accepted as such. I have read somewhere—“A difference that makes no difference is no difference”.

To put it a little differently, I would like to make the following modified statement as—If a psychological test constructed by us in simple Hindi, for use with our population, majority of which consists of illiterates, barely literates, rural and unsophisticated people is better than a well-standardised Western test for use with their populations when used in India, without doing any modifications, it would make a difference to be called a real difference, in the end. On the other hand, if it does not make any difference, I would most humbly be ready to bow my head to be sacrificed.

5 A Start to Meet the Problems Faced

In order to meet the problems faced by us, we started with assessing the extent of the problems one by one and finding solutions that were available, practical and within our resources.

The first problem I faced was the length of the items in the test, particularly in the CMI Questionnaire (Erdmann Jr. et al. 1952). It was used in English with spot translation in Hindi. It contained nearly 200 items referring to different bodily systems and psychological complaints (A–L for physical section and M–R for psychological symptoms, with A–R referring to total complaints). We attempted to translate it in Hindi, so that the same language could be used with every patient. This took care of uniformity of the items, from tester to tester, and over time.

Another related difficulty arose with regard to the total scores. With the spot translations of English version, perhaps our patients could not relate to their problem correctly and did not endorse many items.

Patients' relative positions with regard to the scores remained the same (as shown by the coefficients of correlation between English and Hindi versions remaining high), but the average score significantly increased with the Hindi version. This justified our attempt at developing a Hindi version, as patients could actually feel that the item really described their own difficulties or problems, versus the exact, literal, on the spot translation as attempted earlier.

Another issue was whether so many items were really required or whether the same results could be achieved by fewer items, which were more reliable and valid. So, item analysis was begun, to find out which items were readily accepted (frequently endorsed) by many patients, as opposed to others, which were rarely (if at all) endorsed by non-patients and patients (patients with both medical and psychiatric problems).

Those items that had zero or nearly zero percentage of endorsement were excluded in order to prepare a shorter version of the scale in simple Hindi.

When literature search for item construction was done, it was found that illiterate or barely literate people related better to items when expressed in first person (I, me, my versus you, your). In the English version, as well as in spot-translated Hindi version, the items were asked in the second person only, e.g. every time the items were “do you have” this or that problem. A need was felt for using first-person expressions in these questions. This shortened and simplified the statements (in place of

questions), with the common question at the top asking them to tick the items applicable to them. Thus, the question form was converted into statement form to make it easier, simpler, shorter and quicker, taking less time to complete the test, while it still remained questionnaire (i.e. tick those items that are correct for you).

The “thinking aloud method” was used to make items short, simple and easily comprehensible by our patients. Each item was read aloud, and any difficulties that may be arisen in the patient’s mind were addressed to remove all ambiguity, if any, found in expressions. When required, that ambiguity was removed by modifying the item structure, replacing difficult words by its simplest form, without sacrificing the meaning. This critical self-examination was a necessary evil, which proved quite useful in the end.

A cross-cultural examination was done on items with at least 10 % of endorsements, and system-wise comparisons were done.

It was found that our clinic population endorsed more items relating to gastrointestinal problems and of general weakness, as compared to Western cultures, where the systems involved were cardiovascular and respiratory. Keeping this in mind, item selection was done accordingly. All this research was based on different studies, which were published in professional journals.

Thus, the shortened version of the CMI, modified according our own needs, consisting of 38 items was developed. Section A was physical section consisting of 22 items corresponding to A–L section of the CMI, and divided into Sections A and B. Section B was related to M–R section of the CMI and consisted of mental or psychological items. A provision for a total score consisting of all the 38 items was also. This scale was called as PGI Health Questionnaire N-1 (Verma et al. 1985).

This PGI Health Questionnaire N-1 was presented at one of the annual conferences of the Indian Psychiatric Society and won the prestigious Marfatia Award.

This provided much needed confirmation of rightfulness of our efforts and encouraged us to follow the same line in future also.

6 Continuation of the Work

As the next step, we collected the chief complaints of our patients, describing their complaints, in their own words and in Hindi. This was made possible by Prof Wig’s foresight and direction to all (staff and students of the Department) to write in verbatim the chief complaints in Hindi, even before translating them. I collected all these complaints for 2 years and put them in proper order and saw to it that all types of patients, male and female, rural and urban, of all age groups, of all educational levels and all types of “neurotic” patients were included in this sample. The next step was to put similar complaints together and select the simplest and most comprehensive expression to represent the group. Fifty items were thus selected in this way. This formed the N-Scale, and we defined “neuroticism as a tendency to develop neurotic symptoms when under stress”. Items were numbered 1–50, and the scale was called as PGI HQ N-2. A new scale consisting of 10 lie items was

added to this N-2 scale, after a lot of research with “too-good-to-be-true items” selected from the original EPI Lie scale and the PEN scales of H.J. Eysenck, who had taken them from original MMPI scale of Hathway and Mackinley of Minnesota (Eysenck and Eysenck 1964; Hathaway and McKinley 1940). These items were translated into simple Hindi and used in number of studies using modified instructions to test for different response biases and selecting tendency to measure social desirability response set (or the “Lie scale”). It was used with the neuroticism scale PGI HQ N-2 (Wig and Verma 1979).

In order to distinguish these “too-good-to-be-true” items from neurotic symptoms, these were kept as separate and numbered 51–60. Thus, the PGI HQ N-2 consisted of 60 items in all. Reliabilities and validity of this new scale were established. It was also cross-validated against samples drawn from psychiatric units, from all parts of India. This scale was constructed and standardised and formed subject matter of my PhD, the first PhD of the department.

The two scales N and L of the PGI HQ. N-2 were found to have low correlation, indicating the freedom from the social desirability response set of this neuroticism scale. These N-scores were found to have positive and significant correlations with other measures of neuroticism existing at the time, when used with English-speaking educated samples (who could take tests in both languages). This confirmed that this N-measure had good validity. N-1 and N-2 showed good correlation as expected. Another quality of these tests (N-1 and N-2) was their low difficulty levels as compared to many other personality tests in Hindi. A research paper published on difficulty levels of some personality tests used in India written by me got a good mention in a Presidential Address in Indian Science Congress (psychology section), and this was sent to Professor Wig. It confirmed that we were on the right track and this was being recognised by senior psychologists of the country. By the way, I would like to mention here that this publication (paper on difficulty levels) was published in Indian Journal of Clinical Psychology. I was elected as editor of this journal in 1974 and was re-elected four times (1974–1983). This was all because of our contribution in this area of psychological testing.

7 Measures of Difficulty Levels in Personality Tests

Here, a mention needs to be made, as to how difficulty levels can be measured and what could be the units of its measurement, so that tests consisting of different length of items could be compared. In test of intelligence, it is done easily by comparing the pass percentage of items, where an easy item is one passed by almost all the subjects, and most difficult item as the one that very few or none could pass. In a personality test (there are no right answers to be called as pass), the difficulty levels can be judged in the following ways:

- The number of words used in a test per 100 items (i.e. in percentages calculated) or the number of letters per 100 items (as unit of measurement). It takes time to read each item, to comprehend each item and to respond thereafter. The

assumption being that longer the sentence, it would take relatively longer time and would more difficult. (In fact, we rarely used only one test for each subject e.g. one intelligence test, one questionnaire, one projective test, one memory test and one test for perceptuo-motor function. One had to choose among many tests; hence, one chose short, less difficult tests.)

- The total time taken (greater the time taken, more difficult would be the test).
- The number of difficult words per 100 items. It had to be rated as “difficult” (defined in their own way) by the same group of experts, to be unbiased in rating different tests. This is called the “underlining test”, i.e. underlining “difficult” words.
- Instead of each item making a fresh start, if items are gone through in a quick succession, they would be called easy items. The arrangement of items was such that it made it easy to pass through a lot of items in quick succession, e.g. items, starting with “I”, could be put together, those with “my” in another group and those with “we” together in a third group.
- Tests could be ranked on each of these criteria, so that the lower the rank number, easier it would be. When ranked in this way, we found our own tests were the least difficult ones, on all these criteria, when comparison was made with other tests.

8 Measurement of Response Biases

Another scale constructed by me was for measuring response biases such as the following:

- Tendency to choose the first response choice, in the list of response choices; for example, in yes/no choice list, it is yes, while in the list of no/yes, it is no more frequently selected. This is independent of the question or statement. Surprisingly, it was found that the same person selects the first response, even when the same question is being asked a second time, but in reverse order, hence indicating the response bias.
- Tendency to choose the extreme responses. This can be seen, where, in a list of choices, extreme ones are selected, e.g. either one is extremely happy or, extremely unhappy, never in between. This shows a response bias.
- Tendency to choose the middle most responses. When people do not agree with extreme responses, when they are not sure whether to agree or disagree, they feel safer by avoiding extreme responses and choosing the middle response instead. This shows a feeling of insecurity.
- Tendency to choose either all positive responses or all negative ones. Those well-adjusted, satisfied persons or maladjusted, dissatisfied persons tend to choose such one-sided responses.

Thus, whenever a particular bias is shown in a test, one is confident that the responses of the person do not correspond to the question or statement. In other

words, the responses are independent of the content of the items, hence are untrustworthy and cannot be accepted as true responses.

A scale was prepared where there were no questions or statements—but only responses were given. It is called rightly as “The questionnaire without questions”. Here, there were no random answers; otherwise, all response choices would be equally selected as responses by a group of subjects. Deviation from equal chance probability to a significant extent was considered as a “response tendency” for a person or for a group.

The Lie scale used in the PGI HQN-2 was one such a measure, indicating a tendency to give socially desirable responses. A high lie score, thus, indicated the unreliability of N-score in a given situation, by a specific group of subjects.

This was shown to be the case in a number of experimental conditions, where a significant bias was induced by specific directions (e.g. in interviews for jobs or interviews for a decision for compensation).

Matching of normal and neurotic groups:

The “neurotic” group was selected on the basis of case identification in the psychiatric clinics, while the normal group was selected mainly from the relatives or attendants of these patients. The idea behind this was the assumption that both would be from the same or similar socio-economic background and hence would be well matched. In addition, these groups were also matched on relevant socio-economic variables such as age, sex, education and rural or urban residence. This was necessary for meaningful comparison of test scores.

9 The PGI Memory Scale

My esteemed professional colleague, Dr. Dwarka Pershad, wanted to construct a battery of memory scales for use with our clinic population. He started at first with well-known batteries such as the Wechsler Memory Scales and the Boston Memory Scales.

Both were in English and had foreign norms, but were used in our country with spot translations and with varied success, reported and unreported. The limitations were quite apparent, and a need was soon felt for developing our own memory scales in simple Hindi and to develop our own norms, for use with our clinic population, who were mostly illiterate or barely literate, rural and unsophisticated people.

Selecting various areas of measuring memory, he started experimenting with various subjects which included the following:

- General information and personal data for recent memory.
- Areas for recent and remote past.
- Attention span.
- Memory for words and sentences.
- New learning for new unrelated pair of words.
- Recognition of simple designs.
- Retention for old, familiar words.

He, thus, collected material for his battery of memory scales in 10 such areas and formed a battery called PGI Battery of Memory Scale for clinic population (Pershad and Wig 1977; Pershad and Verma 1990).

All those items were in simple Hindi. Experimenting with different time intervals, he finalised standard times for exposure of stimulus material and the time gap before reproduction/recognition of those stimulus materials. Percentile scores were used, and a profile of scores for differentiating between normal and organic cases was generated.

Finding that obtained scores varied significantly with educational levels, and Dr. Pershad decided to develop separate norms for three educational levels:

- 0–5 years of schooling,
- 6–9 years of schooling and
- Above 10 years of schooling.

This was done for the first time in the field of psychological testing in our country and won recognition for Dr. Pershad and for the department, in the field of psychiatry in India. This battery of memory scales was for his PhD work under the guidance of Prof. Wig. This was another major contribution of the department in the field of psychological testing.

10 A Beginning of Teamwork

We (I and Dr. Dwarka Pershad) worked together for the construction of reliable and valid tools for our clinic population, and significant among them were as follows:

- Modified shorter version of Bhatias' battery of performance tests of intelligence (B.B.P.T.). In this pioneering work of Dr. C.M. Bhatia, in the construction of the battery of performance tests, there were many significant contributions such as the following:
 - Separate norms for illiterates—used for the first time in India.
 - Large sample of students, 11–16 years of age, from different areas of Uttar Pradesh.
 - It consisted of 5 subjects—viz. Kohs' Block Design Test, Alexanders Pass—Along Test, Digit Span Test (for literates but alphabets for illiterates), Picture Completion Test and Picture Construction Tests.

We found some good points in it, but also some drawbacks:

- Test norms were more than 15 years old and needed revision.
- Scoring was limited to 1 min or 2 min only, requiring wider range of scoring for greater reliability.
- Extending norms for adult groups also. Means and standard deviations were computed.
- Norms for three educational groups and for different age levels.

When we approached Dr. C.M. Bhatia, he readily agreed for us to carry out the necessary modifications, which we did and published our results in the form of a practice manual. We used deviant IQ directly calculated from the raw scores.

- WAIS-R verbal scales in Hindi. After obtaining the necessary permission from the author and publisher, we selected few verbal scales such as Information, Digit Span, Arithmetic and Comprehension for translation in Hindi and administered them to normal population in different age and education groups. Data were collected from other centres in India such as Agra, Baroda, Bikaner and Ahmadabad with the help of co-investigators there. After calculating means and standard deviations for each education groups, deviant IQs were obtained directly from the raw scores.

This was the deviation from the original method used. This information was given in the manual concerned.

- PGI Battery of Brain Dysfunctions. It consisted of a number of subjects such as the PGI Memory Scale, Bhatias' Short Scale of performance test of intelligence and WAIS-R verbal scales in Hindi. The information about its construction, administration, standardisation reliability, validity and normative data is given in the manual published. Measures of visuo-motor functions were also added after doing research on the BVMG test, the Nahar and Benson scale, and the Hoopers Visual Organization Test (VOT). Dysfunctional Analysis Questionnaire (D.A.Q.) was also prepared to assess the deviation from previous functioning (i.e. before the illness started). Norms for the Standard Progressive Matrices were also established for different age and education groups using deviant IQs (Pershad and Verma 1990).

11 Other Scales or Tests Developed

A number of new psychological tools were prepared and used for research carried out for PhD degrees in the department (including those of other departments where guidance was provided by clinical psychologist of our department).

These include the following:

- A schedule for measuring temperament in children (Savita Malhotra).
- A childhood psychopathology measurement schedule (Savita Malhotra).
- Parental handling questionnaire (Savita Malhotra).
- Psychoticism scale in Hindi (Manju Mehta).
- PGI locus of control scale (D.K. Menon).
- PGI achievement value India (D.K. Menon).
- A shorter way to assess the IQ (Adarsh Kohli).
- Scale for assessment of attitude towards yoga (Poonam Grover).
- Medication attitude scale (Poonam Grover).
- Maternal attitude and maternal adjustment scale (Karobi Das).
- A scale for personality disorders (Pratap Sharan).

- A coping checklist (Ritu Nehra).
- Social support questionnaire (Ritu Nehra).
- Paediatric development screening test (Amita Puri).
- Home based-programme for MR children (Mala Seshadri).

Other psychological tests constructed or used in revised forms for MD thesis, for funded (from WHO, I.C.M.R., I.C.S.S.R. and PGI research schemes) and non-funded departmental projects include the following:

- PGI General Well-Being Scale.
- PGI Quality of Life Scale.
- PGI Job Satisfaction Scale.
- PGI Battery of Assessment of Mental Efficiency in the Elderly.
- Hindi Personality Trait Inventory.
- Bond's Defence Style Questionnaire.
- Learned Helplessness Scale.
- Hope Scale.
- Pleasure Scale.
- Sensation Seeking Scale.
- SPD Scale.
- Sex Knowledge and Attitude Scale.
- Geriatric Depression Scale.
- Eating Attitude Scale.
- Psychiatric Disability scale.
- Relapse Precipitant Inventory.
- Self-efficacy Scale.
- Alienation Scale.
- Cognitive Behaviour Scale.
- Vignettes for Attitudinal Research.
- Hindi MPI Scale.
- Authoritarian Scale.
- Tests for Linguistic Competence.
- Illness Behaviour Questionnaire.
- PGI sentence completion tests—Forms G (General)
 - Form M (Married individuals).
 - Form S (Students).
- Social Burden Scale.
- Family Burden Scale.
- Home Environment Scale.

Projective tests:

A number of projective tests were devised/used, and normative studies conducted on them:

- Draw-a-Person Test—Body Image Concept of Students.
- Body Image Disturbances of Psychiatric Patients.

- Body Image Disturbances of Patients Undergoing Cardiac Surgery.
- Sack's Sentence Completion Test (Hindi).
- Indian Adaptation of T.A.T. (Story-making test).
- Rosenzweig Picture—Frustration Study (Indian adaptation).
 - For adults.
 - For children.
- For studying attitudes of patients:
 - Standard stories were used (vignettes) depicting various illnesses.
 - Standard pictures depicting various psychiatric disorders.

Both these were constructed and used in the department.

- Word Association Test.
- Rorschach Test—(Indian Norms, modified instructions).
- Somatic Inkblot Series.
 - Booklet Form.
 - Card Form.

Indian norms and reliability studies were devised for the following:

- Three Wishes Test.
- Projective Question—(If reborn—what animal you would like to be—and why).
- Draw a Person—Practice Manual and Norms.

12 Manuals

Construction and standardisation of psychological tests is one thing, writing useful and practical manuals, complete in all respects, making it ready for use by other researchers, throughout the country is equally important. This was taken care of by us working in the department. The details are as follows:

Publishers:

A. National Psychological Corporation, Agra.

- Manual for PGI Health Questionnaire N-1 in Hindi (1985).
- Manual for measurement of dysfunction and Dysfunction Analysis Questionnaire (DAQ) in Hindi (1985).
- Abridged manual of CMI Health Questionnaire in Hindi (1983).
- Manual for Bhatia's Battery of Performance Tests of Intelligence (Short Scale) (1988).
- Concept and measurement of Intelligence (1988).
- Manual for Malhotras' Temperament Schedule (1988).
- PGI Battery of Brian Dysfunction (PGIBBD) (1990).
- Manual for PGI Battery for Assessment of Mental Efficiency in the elderly (PGI-AMEE) (1996).

- Draw a Person Tests Manual.
- Manual for ADHD.

B. Ankur Psychological Agency, Agra.

- Manual for PGI General Well-Being scale (Verma and Verma 1989).

C. Rupa Psychological Corporation, Varanasi.

- Manual for Hindi PEN Inventory (1988).
- Manual for PGI Locus of Control Scale (1988).
- Manual for PGI Achievement Value Index (1988).

An ideal test manual is expected to include standard stimulus content, ways to administer the test, way to score objectively, specific, local normative data and standardised method of interpretation, with established reliabilities and validities. All these should be open for inspection and acceptance or rejection. The concept itself has to be well-defined, meaningful and acceptable for critical evaluation by others, if they so desire and when they decide.

As such, things require a great deal of time and effort, and this is difficult for research workers working with limited resources available. This, however, is not an acceptable excuse for poor quality of products and for avoiding criticism. All criticism is healthy and motivates us to do more and better and hence should be welcomed with an open heart.

13 Science and Religion in Measurements

This brings us to look philosophically at the issue of relation between the science and religion and how this can help us in our work. There are many individual facts to be discovered. We cannot limit ourselves to discover facts alone, such as constructing psychological tools to find more and more scientific facts. It is rightly said that “those who refuse to go beyond facts rarely go as far as facts”. Facts have to be linked together in a meaningful way and used for a purpose.

Whereas scientific methods help in collection of newer facts and tools to measure something, we have to pursue the search for their meaning and application religiously, like a determined devotee, trying to please the “God of Measurement”.

In Hindu Philosophy, there are many Gods such as *Brahma* (for generation/creation), *Vishnu* (for protection), *Mahesh* (for destruction), Vayu Devta (for air), Agni Devta (for fire), Indra Devta (for rain) and *Yamraj* (for death). I may therefore be forgiven for introducing another god, the “God of Measurement”, and I am willing to submit myself, all my work on measurement and construction for psychological tests to him. Each test can be visualised as flowers or fruits for offering to get his blessings.

The method is scientific, while the pursuit is religious, i.e. blindly and faithfully, with all our forces and resources applied to get scientific results. Each tool is like another *Brahmastra*, i.e. an instrument or weapon to get “good” results. The

weapon should be prepared in a scientific way. The concept has to be clearly and unambiguously defined and religiously followed (with approved *mantras*) to get effective results (to achieve measurement of the concept). We are ready to face others' weapons (criticisms, challenges) with our *Brahmastra* (test scores). Both science and religion are involved. But for the collection of scientific facts, collected religiously, we would still believe that the earth is flat (not round), that sun moves around earth (not earth revolving around the sun), and that an atom is indivisible (not divisible into nucleus and sub-atomic particles).

The same thing is true about newer methods of measurements, and depending upon the newer facts discovered, we modify our tools, e.g. if difficulty levels of language create problems for those not properly educated, we use simpler language and prepare tests that can be used with all education levels, with local norms for different educational groups.

In the West, when Eysenck found the MPI was difficult for lower educational levels, a new test with simpler language, the EPI was constructed (Eysenck and Eysenck 1964), and when Cattell found that 16PF Test Forms A and B were posing problem with education levels, he constructed forms C and D, and even E and F making the language simpler (Cattell 1989).

Surprisingly, our so-called experts (university professors, used to working with highly educated people) denounced any such notion of difficulty levels and still preferred to use English tests as such, advocating their use with our clinic patients as well. We refused to accept the suggestions of "authorities", i.e. senior university professors, as science does not accept any authority, as science starts with doubts and ends with the valid proofs. Religion starts with only beliefs and refuses to accept doubts at any level. We went on using scientific methods religiously, refusing to bow down to the so-called authorities.

While future psychologists working in the department are free to work according to their own faith, in their own selected areas, to realise their own dreams, I would like to again point out my own rebellious attitude, against established authorities, in our more example as another attempt to justify my approach. It is a case of the Seguin Form Board Test. The original norms were devised by Cattell (in the Guide to Mental Testing), based on research done abroad. We were using those norms. A funded research, on large samples (over 500) in each of the three language groups in South India, published in a book form, suggested that children in our country were 20 % less efficient as found in their research. This was taken up as a challenge, and we found many facts that showed poor validity of their claims:

- This test was used by school teachers (not trained clinical psychologists), from school situated on roadsides, with lot of disturbance from continuous traffic (not in a psychology laboratory with no disturbance), and from class rooms, in the presence of other students.
- Standard procedures were not followed. For example, no practice trials were given before starting data collection. Use of preferred hand (right or left) was not strictly followed. Sometimes even both hands were used.
- Age of the children was not reliable (or reported in the book itself by the authors). Many children had the same date of birth filled up by school staff to

enable students get admission without any problem of age. Parents were not sure in most cases about the exact ages.

- Either no stop watch was used and/or the wrist watches were used in wrong manner.
- To “start” spoken aloud was to be taken as actual start of the work and not necessarily when the child actually starts. Most of the children were from rural background, not used to competition and did not realise the importance of doing the work as fast as possible.
- To “stop” was not noted, when the child had actually completed the task, but only when he/she verbally indicated. Many children were not sure when to stop and say I have done the work.

As this is a speed test of intelligence, the above factors had a significant role in influencing the test scores.

It was more a test of somehow compiling the data and not necessarily a test of real intelligence. Role of non-cognitive factors played a part, and results were wrongly interpreted as real slowness of Indian children. All these facts were reported in a paper—“Are Indian children slow?” published in the *Child Psychiatry Quarterly* (Verma et al. 1980).

This only shows that a common test does not necessarily mean the same test, if not used in a specified (standard) manner, as described in the test manual.

It also shows the importance of publishing test manuals in addition to test construction (to avoid wrong or hasty conclusions). Finally, the importance of religiously following test manuals for deriving or correcting interpretations scientifically was also demonstrated.

14 Concluding Remarks

It is good writing about my first and foremost love—i.e. psychological tests, in the area of measurement. People write their autobiographies to justify their past mistakes and leave it for others to give their judgments. So, writing about our psychological tests in use, I surrender myself before the God of Measurement and give ample opportunity to people who matter and time to give their thoughtful judgments about the quality, reasons and outputs of our department.

I was allowed to work with full devotion to achieve my aims and fulfil my dreams, for which I am indeed very grateful to the God of Measurement, to the Head of the Department, my seniors, colleagues and students of the department. Work is its own reward. There can be no greater reward than the satisfaction of doing ones’ work honestly and sincerely, and if any mistakes are made in this process, I take the full responsibility for it.

Let those errors be recognised and corrective measures taken to improve them further. This is my hope and wish.

The journey towards making perfect psychological tests is an endless pilgrimage to reach the vague, undefined, abstract concept of God of Measurement. While

the task is difficult, if not impossible, the journey is pleasurable for all the selfless devotees offering their lifetime work to it. So let us all enjoy. The journey continues. The procession is on. Like the normal probability curve, this curve approaches the baseline, but never touches the baseline. There is always a scope for hope, for better endings, by better or smarter people. After all, "God is in the Heaven; everything is right with the world, with the department and with the movement of psychological tests".

References

- Bhatia, C. M. (1955). *Performance intelligence test battery*. Bombay: Oxford University Press.
- Cattell, H. B. (1989). *The 16PF: Personality in depth*. Champaign, IL: Institute for Personality and Ability Testing Inc.
- Erdmann, A. J., Jr, Brodman, K., Lorge, I., & Wolff, H. G. (1952). Cornell Medical Index-Health Questionnaire: v. Outpatient admitting department of a general hospital. *Journal of the American Medical Association*, 149(6), 550–551.
- Eysenck, S. B. G., & Eysenck, H. J. (1964). An improved short questionnaire for the measurement of extraversion and neuroticism. *Life Sciences*, 305, 1103–1109.
- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, 10, 249–254.
- Pershad, D., & Wig, N. N. (1977). *The construction and standardization of clinical test of memory in simple Hindi*. Agra: National Psychological Corporation.
- Pershad, D., & Verma, S. K. (1990). *Handbook of PGI battery of brain dysfunction*. Agra: National Psychological Corporation.
- Raven, J. C., Court, J. H., & Raven, J. (1998). *Manual for Raven's standard progressive matrices*. Oxford, England: Oxford Psychologists Press.
- Verma, S. K., Pershad, D., & Randhawa, A. (1980). Are Indian children slow? Report on enquiry with a speed measure of intelligence. *Child Psychiatry Quarterly*, 13(1), 67–71.
- Verma, S. K., Wig, N. N., & Pershad, D. (1985). *Manual for PGI Health Questionnaire, N-1, a simple neuroticism scale in India*. Agra: National Psychological Corporation.
- Verma, S. K., & Verma, A. (1989). *Manual for PGI general well-being measure*. Lucknow: Ankur Psychological Agency.
- Verma, S. K., Pershad, D., & Mahajan, A. (1990). Hindi translation of Personality Trait Inventory. *The Creative Psychologist*, 2(1), 23–26.
- Wig, N. N., & Verma, S. K. (1979). *Construction and standardization of PGI Health Questionnaire N2*. Agra: Agra Psychological Research Cell.

Developments in Psychiatry in India
Clinical, Research and Policy Perspectives
Malhotra, S.; Chakrabarti, S. (Eds.)
2015, XV, 636 p. 1 illus., Hardcover
ISBN: 978-81-322-1673-5