PART 3. GATE NORMS FOR SPECIFIC OCCUPATIONS
Personnel selection tests in industry are primarily used to select in an objective manner potentially successful candidates. This prediction is necessary because on the one hand different jobs need different kinds and quantum of abilities and on the other hand, individuals are endowed these abilities in varying degrees.

Major Validation Designs and Basic Selection Model

The accuracy of prediction depends upon how properly the test for selection are "validated". The concept of validity is the extent to which a test is capable of achieving its aim or purpose. Since the objective of the test development study is to develop a set of norms which will predict success in a specific occupation, careful attention should be given, before a study is conducted, to formulating an experimental design for test validation which will best accomplish this objective. The two most widely used experimental designs for test validation are the concurrent validation design and the longitudinal validation design.

Concurrent Validation Design: The chief characteristic of this design is that the battery of predictors and the criteria are administered simultaneously to people already working. The validation and cross-validation are conducted on current employees.
Longitudinal Validation Design: The chief characteristic of this design is that the job proficiency of the sample is not known when the tests are administered. Thus data on criterion cannot be obtained until later. The standardization sample usually consists of applicants about to take up a job, and followed up later for their success on the job.

These validation procedures involve certain similar steps and procedures which are described.

a. Determining the Job Requirements: Through job analysis those aspects which are necessary for effective performance should be identified and isolated, and their relative value and importance should be established. A thorough knowledge and understanding of the job is supremely important and must precede the use of any test in the selection and placement of workers.

b. Selecting Predictor: A battery of tests to be tried to the extent to which it can measure attributes deemed important for success on the job. Information on job analysis can serve as a basis for selection of tests.

c. Selecting Criterion: An indicator or criterion to measure the extent to which a worker is successful should be evolved. The criteria for good performance in an industrial setting can include the worker's merit rating, or such clearly defined variables as absence, tardiness, rate of productivity, waste, accidents, and grievances. Any criterion must, of course, reflect the important inputs for any activity, and
must be reliable, relevant, and free from biases.

d. Measuring Performance on Predictors: The next step is administering predictors to a large group of workers, by giving the predictor to present employees or by giving it to new recruits.

e. Obtaining Information on Criterion: Criterion information must be obtained along with information on predictors (concurrent validation) or after a specified time (longitudinal validation).

f. Relating Predictor to Criterion: Whether a true and meaningful relation exists between the predictor scores and the criterion must be determined. Only if such a relationship exists can the selection process be successful. This involves a statistical process where several methods can be used, the commonly used method being correlational.

g. Deciding the Utility of the Selection Device: A decision to be reached on using the predictor for new recruits. This will be conditioned by the size of correlation and its significance, the number of applicants, the number of job openings, the proportion of present employees considered successful, and the respective variance of the successful and unsuccessful workers.

h. Cross-Validating the Predictors: Using a different group of candidates, determine if the original relationship between predictor and criterion variables hold. If it does not, the relationship originally
found was probably due to chance factors and, therefore, cannot be relied upon.

Comparison of Validation Designs

Each validation design has certain advantages and disadvantages. The advantage of the concurrent validation design is that results can be obtained fast, the test administration is economical since the number of testees is known, and the criterion data can be collected simultaneously.

However, this design has several practical disadvantages, the key one being that employees are very often reluctant to take workers off the job for administering the test. A technical disadvantage is that the present employees on the job may represent a highly select group, since those not satisfied with the job had left on their own. Such restriction in the range of talent reduces the magnitude of correlations obtained between test scores and the criterion.

The longitudinal validation design has several technical advantages over the concurrent validation design. The standardization group is not seriously affected by the restrictions in the criterion and test scores. Hence the problem of demonstrating relationships between test scores and criterion scores is somewhat eased.

Another advantage is that the experimental sample is drawn from the same population (applicants) on whom the test will be used for operating
purposes. In addition, this design rules out the possibility of the 
scores being influenced by specific training and experience on the job 
since only fresh recruits are tested.

However, the most serious practical disadvantage of the longitudinal 
design is that it takes considerable time to complete the studies and 
obtain the results. Quite often, it does not result in any proved tests 
until the programme has been underway for several months or several years. 
In some cases, applicants who were tested do not remain on the job long 

enough to yield suitable criterion data.

Test Validity and Decision Theory

The validity coefficient is usually defined as the correlation of test 
scores with the outcome or criterion score. How high should a validity 
coefficient be? Although this question has no general answer, everybody 
agrees that the obtained correlation should be high enough to be statisti-
cally significant at some acceptable level, such as the .01 or .05. How-
ever, after obtaining a significant correlation between test scores and 
criterion, the size of the correlation against the uses to be made of the 
test should be evaluated.

The most common method used to evaluate the utility of the test has 
been to translate the validity coefficient into the "index of forecasting 
efficiency", defined as the percentage reduction in errors of prediction 
by reason of correlation between two variables. The general simplified
formula is \( E = 100 \left( 1 - \sqrt{1 - r^2} \right) \). It will be noted that if the validity were perfect, the index of forecasting efficiency could be 100. Whereas, with a test having zero validity, the index of forecasting efficiency would be zero. Under these conditions, the prediction is no better than a conjecture. Between these two extremes lie the errors of estimate corresponding to tests of varying validity.

Another mode of interpretation of validity is in terms of \( R^2 \), the coefficient of determination. The coefficient \( R^2 \) gives, when multiplied by 100, the percentage of variance in \( Y \) that is associated with or determined by variance in \( X \). When \( r = 0.50 \), the percentage of the variance in \( Y \) that is accounted for by variance in \( X \) is 25 or one-fourth.

According to traditional interpretations based on these formulations, a high validity coefficient is necessary to benefit substantially from testing. The index of forecasting efficiency describes a test correlating .50 with the criterion as predicting only 13 percent better than chance; the coefficient of determination describes the same test as accounting for 25 percent of the variance in outcome. Even if the validity coefficient is as high as .80, the index of forecasting efficiency and the coefficient of determination are only 40 percent and 64 percent. In other words, even a test of such high validity predicts the individual's criterion performance with a margin of error that is only 40 percent lower than a conjecture.
Ghiselli's (1955) comprehensive review of published and unpublished studies showed the average validities ranging in the 30s and low 40s. If this is the position of the validity coefficients, the outlook is rather discouraging.

Recently Cronbach and Gleser (1957) have dealt with the use of psychological tests in making personnel decisions in the context of decision theory, advocating an approach that provides a more meaningful criterion for evaluating a psychological test used for selection and placement. It is an attempt to put the decision making process into mathematical form so that the available information can be used to make the most effective decisions under specified circumstances. "The value of a test", contend Cronbach and Gleser, can be stated in terms of the specific type of decision problems, the strategy employed, the evaluation attached to the outcome, and the cost of testing" (Cronbach and Gleser, 1957, p. 32).

In recognizing first the inadequacies in such measures as $E$ and $R^2$, Taylor and Russell became the precursors of the present decision theory in psychological testing. They examined what proportion of employees were likely to be "satisfactory" before and after selection by means of tests. This proportion depends on the selection ratio, the validity coefficient of the test, and the proportion of successful applicants selected without tests. An increase in proportion of successful applicants is an index of the tests effectiveness. A test with validity .30 permits substantial...
improvement when the selection ratio is low, and under some conditions a
test with validity .80 gives a perfect selection. Thus according to the
decision theory, the value of test information should be judged by the
degree of improvement in decisions over the best possible decisions made
without the test, whereas the conventional validity coefficient reports
how much better test decisions are than chance decisions.

Following the work of Taylor and Russell, Brogden (1946) evaluated deci­sions directly on the utility scale to interpret the validity coefficient and concluded that the expected increase in output is directly propor­tional to the validity, regardless of the selection ratio. He argues
that a test of validity .50 effects 50 percent of the improvement that
would result from using a perfect test. Such an evaluation of test vali­dity is obviously much more favourable than that based on the error of
estimate.

Another aspect of the decision theory is the evaluation of outcomes. The
theory calls for formulating a strategy which maximizes the average gain
over many similar decisions. The expected pay-off from an individual in
a particular category can be determined by weighing the value of each
outcome that might result from him by its probability and then summing
over all outcomes. The cost of the procedure for gathering information
on which the decision is made can be expressed in utility units and
then deducted from the expected pay-off.
Perhaps the most far reaching feature of the decision theory is that it takes a more favourable view of tests with low validity. Traditionally, such tests are considered as poor predictors, but these become valuable when selection ratios are low.

The examples cited provide a few glimpses into ways in which the application of the decision theory can affect the interpretation of test validity. "Conventional test theory assumes that we use tests to obtain numerical measures on an interval scale, as in the physical sciences. That is rarely or never true. The function of psychological and educational tests is to aid in making discrete decisions. The greater contribution of decision theory is to help testers see this function more clearly" (Cronbach, 1966, p. 58).

Moderator Variable

A recent development of significance in the interpretation of test validity is the use of moderator variables. A moderator variable can be described as one which, when varied systematically, has an effect upon the magnitude of relationship between two or more variables (Blum and Naylor, 1968). The traditional validity model assumed that prediction errors characterized the test rather than the person, and that these errors were randomly distributed among persons.

It has been recognized that this approach was not altogether correct. The moderator variable approach emphasizes the interaction between tests...
and persons. Such an interaction implies that the validity of a given test may vary among sub-groups of individuals or even among individuals in a population. For example, a test may be a better predictor for men than for women or for younger than for older applicants. In these examples, sex and age are moderator variables (Saunders, 1956).

The validity coefficient may be low when computed for the total group. But if the group is sub-classified according to some identifiable characteristics, the validity may be high in one group and negligible in another. Thus the test can now be used effectively for the first group but not for the second. Interest and motivation often function as moderator variables in individual cases. Other common moderator variables are sex, age, education and socio-economic status. The predictive validity of tests can be improved if the moderator variables are incorporated in the appropriate selection strategy. The example of such a strategy can be found in Dunnette (1966).

Multiple Cut-off Scores V. Multiple Regression

When a person's standing on one variable, usually designated as a criterion, is to be predicted based on his/her position on several other variables, usually called predictors, the most commonly used procedure involves (1) assuming that the relationships involved are linear and (2) employing the least squared error multiple regression weights. This has been the traditional method of expressing test validity and its weaknesses were enumerated in the earlier section. Moreover, since $R$ and $R^2$ are a summary
statistics involving scores across the entire range, wrong interpretations can arise if assumptions concerning bivariate linearity are not fulfilled.

Another objection raised against the multiple regression method centres around the question of compensatory qualification. It is possible for someone who is weak in one area to compensate by being strong in another. It is possible, however, that certain types of activity may require essential skills for which there is no substitute. For example, a pilot must have good eyesight and good theoretical training and good coordinating ability. He cannot compensate for poor eyesight no matter how good is his theoretical training or coordinating ability might be.

An alternative strategy for combining test scores utilize multiple cut-off points. This procedure involves the establishment of a critical or minimum qualifying score on each test. An individual is considered qualified only if he falls above such a minimum score on each of the tests. It does not allow for compensation. An example of this technique is provided by the research conducted on GATB by U.S. Department of Labour.

The multiple cut-off is more stable from sample to sample. Because the samples usually available for test development studies are relatively small, the weights which result from the application of the multiple regression method can fluctuate widely. Moreover, assumption of any particular kind of distribution or regression line is not made in this method.

Another advantage of the multiple cut-off technique is that it can be more easily explained to the layman than the multiple regression technique.