The word intelligence conveys different meaning to different situations or environments. Many psychologists have given the proper meaning of the word intelligence. Tewari (1987)* has clearly written the conceptual meaning that intelligence is closely related to intellect which includes observing, thinking, understanding, remembering, and all ways of knowing. William A. and Lehmann (1984), ** have mentioned that the terms 'Aptitude', 'Ability', 'Intelligence' and achievements are used interchangeably by some, while others suggest that subtle shades of meaning distinguishes them. The first three terms are usually considered to have the most meaning in common. According to Frances Galton, (Lewis 1974), *** Latin term 'Intelligence' is to refer to individual differences in mental ability.

* Dr. (Mrs.) Rama Tewari. General Mental Ability Test (GMAT). Agra University Agra 1987, p.2.


No project or research can be started without a definite definition of the work. While discussing the issues and problems of psychological testing, Sansanwal (Dr. M. C. Jain et al., NCERT),* emphasised that the test constructors are required to give operational definition of the variable for which the test is constructed. But this aspect is neglected by majority of researchers. This makes the interpretation of the scores difficult. Along with the definition, the aspects tested should also be given.

Thus knowing the importance of the definitions, many people had given number of definitions on the intelligence and intelligence test. They had given different definitions of their own but with a common focus. In our everyday life we commonly and informally talk about intelligence of the people. But no one can simply categorise through informal judgement without a test scale. Although the judgement of the common people are acceptable to many, but the psychologist do not agree to those suggestions. Many writers have given a lot of definition on intelligence and intelligence test.

Norman E. Gronlund (1981)* advocated that the tests designed to measure learning ability have traditionally been called intelligence tests. This terminology is still used for many individually administered tests and for some group tests, but its use is declining. But in place of intelligence tests have come some such terms as mental ability tests, school ability tests, cognitive tests and scholastic aptitude tests.

As he was trying to define intelligence tests, Lewis R. Aiken (1971)** reminded to remember that so called intelligence tests are actually tests of achievements. The items on intelligence tests represent attempts to assess individual differences in the effects of experiences that are common to nearly in the culture.

Any type of test requires a clear-cut purpose. But to visualise the purpose of a new test is not an easy task indeed. The purpose may be used as an indicator of educational and vocational effectiveness in future. On the other hand can be developed for the purpose of career planning, guidance, classification and decision-making.

** Lewis, op. cit., p. 113.
Some of the authors also had identify the purpose of psychological testing and assessment in their own language pattern. As proposed by S. Venkatasen (Dr. M.C. Jain et al. 1992),* the purpose of psychological assessment varies from screening, identification, classification, placement, programming to certification.

In his test manual, Mehrotra (1972),** explained that India has adopted the practice of diverting pupils to various types of courses after classes VIII and X. The orientation of the pattern of education by making it multipurpose at the higher secondary stage and the diversification of courses including occupational courses at that stage has doubtless offered bright opportunities to young pupils in schools but it has also created a number of problems. Psychologists have no ready-made solutions to those problems but they always felt the need of some tools and techniques to enable them to study, predict and modify human behaviour.

As suggested by A. Anastasi (1976),*** intelligence tests should not be used to label an individual but to help in understanding him. And he further point

* Dr. M.C. Jain, et al., op. cit., p. 84.
** P. N. Mehrotra. Mixed Type Group Test of Intelligence for children from 11 to 17 years. Agra University, Agra, 1972, p.
pointed out that the information provided by tests is
being used increasingly to assess individuals in their
educational and vocational planning and in making
decisions about their own lives.

History of psychological testing and test
scale development shows that the psychological branch
used group scales for classification of soldiers mental
ability shortly after the U.S. entered the World War I.
Since then a number of non-verbal group scales had been
systematically designed. It was written in his test
manual by Jalota (1976), * considering from different
angles, it can be accepted that the group test scales
are better than individual performance scales.

It has been mentioned in their book by
William and Lehmann (1984), ** that group aptitude tests
are far more extensively administered intelligence tests.
They are much less expensive and generally give results
comparable to those of the more time-consuming
individual tests.

* S. Jalota's Group Test of General Mental Ability.
  Manual of Revision, Published in 1972.
According to Hermon Nelson (Lewis 1971), the items are arranged in spiral omnibus forms; that is, several different types of items are distributed the scale instead of being grouped by sub-test.

While explaining about the omnibus type Lewis (1971), wrote that in the spiral omnibus formal, the several types of items comprising the test are mixed together and arranged in order of increasing difficulty. Examples of this scale are listed:

The Army General Classification (Alpha),
The Otis-Lennon Mental Ability Tests,
The Hermon-Nelson Test of Mental Ability.

Non-verbal test items are consisting of pictorial, diagrametic and geometric figures. These items measures the perceptual and conceptual ability level of the testees without any prior training to discriminate amongst them. As a support, Anne Anastasi (1976), Non-verbal battery sub-tests use no words or numbers, but only geometric or figural elements, the items bears reliability little relation to formal school instruction.

** Ibid., p.130.
According to Freeman (1965),* Authors of non-verbal tests of mental ability seek to measure the same mental process tested by means of verbal scales. They hold that the problems presented in diagrams, pictures, charts and geometric forms closely parallel to those presented by means of language and number. Language and numbers are symbolic systems that represent something else for example objects, qualities, events and actions. With use of language and number he is able to analyse, synthesise, classify and organize his perception. Therefore they prefer to test intelligence by means of verbal and numerical materials. However, they would use non-verbal tests when these are made necessary by developmental immaturity or or language or cultural handicap, to gain the insight that these tests provide if they are adequately scaled in difficulty.

Non-verbal items with instructions can be used when there is linguistic handicappness, but has a common

link language and the respondents are literate. Some of the non-verbal tests have been constructed with different sub-tests forming a particular battery.

Lyman (1978),* so called non-verbal tests contain certain no verbal items; however, words almost always used in the directions. Some authors prefer to use non-language to describe tests that have no verbal items but for which the directions are given either orally or in writing.

In the opinion of Anastasi (1975),** a non-verbal battery may have the sub-tests of figure classification, figure analogies, figure synthesis. Ibid (1976),** in the same book, Figure 46, Item Illustrative of the Otis-Lennon Mental Ability Test, every item has given instruction as

(a) Classification: Mark the picture that shows a flame, does not belong with the others.
(b) Verbal classification: Mark the picture that shows a flame.
(c) Quantitative reasoning: Mark the picture that shows the same number of dots as there are parts in the circle.

---


*** Ibid, p. 309.
Again Anastasi (1976)*, Figure 41, Sample Items of Culture Fair Intelligence Test Scale 2 by R.B. Cattel,

Test I. Series: Mark the items that completes the series.

Test 2. Classification: Mark one item in each row that does not belong with others.

Test 3. Matrices: Mark the items that correctly completes the given matrix.

Non-verbal items require a lot of thinking. Unless the items are constructed adequately, no result can be expected as which the test was meant. Some tests may be culture-free, some other may be culture specific.

In this context Anne Anastasi (1976)** clearly pointed that no single test can be universally applicable or equally 'fair' to all cultures. There are as varieties of culture-fair as there are parameters in which cultures differ. Non reading test may be culture-fair in one situation, a non-language test in another a performance test in a third.

Lewis (1971),*** suggested that it is probably impossible to construct an intelligence test that is independent of experiences varying from culture to culture.

---

* Ibid., p. 290.
** Ibid., pp. 345-6.
Bhide (Dr. M. C. Jain et al.), *according to me nothing in the world is 100% 'culture-free.' The concept of culture free can be perhaps applied 100% to robots but human beings are not robots. Therefore every utterance, every thought of human being is so coloured by his/her experiential background and the society in which he/she has grown. So the psychological tools should be as much as possible composite.

To develop a good and valid test, foremost requirement is to choose a truly representative standardization sample. The standardization sample may be made truly representative only when the sufficient number of units are selected from all strata and levels of the population.

In the words of Gupta (1987), **Random sampling does not mean haphazard selection or selection of sample units as they occur to the investigator. But any random sample size 'n' drawn from any population must have the same probability of being selected. Secondly when a population is heterogeneous or different segments or strata in the population, then it is stratified.** So we stratified

---

the population by dividing it into strata so that each stratum is more or less homogenous, and make a random selection of sample from each stratum.

According to Goode and Hall (Hans Raj, 1979), while stratified sampling is placed here in distinction to random sampling, this does not mean that it does not employ randomness. Young (Hans Raj, 1979), the main objective in stratification is to secure a more reliable sample. Sometimes the gains in stratification may be very high and at other times very trivid.

Scoring of respondents marks is one of the major part of test construction and standardization. Scoring number of responses without a proper plan is so difficult, time consuming and liable to error through wrong transferring of the data.

An expert in the field Lewis (1971), also suggested that a strip key or stencil for hand-scoring of test booklets or answer sheets can be prepared quite easily. A strip of cardboard containing the correct answers at positions corresponding to the spaces in the test booklet where answers are to be written makes an


adequate strip key. A scoring stencil for use with a special answer sheet can be prepared from a blank sheet of paper or cardboard by punching out the spaces where the correct answers should appear.

Item analysis is the most important function in the test construction. It is the process of examining the responses to each test item so as to judge its own quality. According to William and Lehmann (1984),* there are more than 50 different item-analysis procedures. They suggested to follow two methods of calculation for item analysis. For each item, compute the percentage of students who get the item correct. This is called the item difficulty index (p), which can range from .00 to 1.00. Formula for item difficulty is: Difficulty = \( \frac{R}{T} \times 100 \). Compute the item discrimination index for each item by subtracting the number of students in the lower group, who answered the item correctly from the number in the upper group who got the item right, and by dividing by the number of students in either group, by the given formula

\[
\text{Discrimination} = \frac{Ru - Rl}{(1/2) T}
\]

As stating they warned that if one did divide the total group into two halves but put the top 27 percent upper

and the bottom 27 percent in the lower group, one could obtain an estimate of item difficulty by dividing the number of persons in the two groups who answered the items correctly by the total number of people in those two groups. (Be careful! Not the total number of students tested!).

Norman E. Gronlund (1981),* suggested the formulii for computation of difficulty index and the discrimination index. The procedures recommended for computations were as following:

1. Score all the answers,
2. Arrange the test papers from highest to the lowest scores.
3. Make two groups 27% upper with the total scores and bottom 27% with the lowest total scores.
4. Then remaining middle papers to be placed aside and not to be used in the analysis.
5. Compute difficulty value and discrimination power by the formulii,

\[
\text{Difficulty} = \frac{R}{T} \times 100
\]

\[
\text{Discrimination} = \frac{R_u - R_l}{1/2 T}
\]

According to Lindeman (1967), three kinds of information can be obtained by analysing the responses to an individual test item (1) the general difficulty level of the item (2) its discriminatory power and (3) the response pattern of the item. In a test some easy items should be included in order to encouraged the students of low ability and some difficult items should be included to challenge the abler students. And for statistical reasons, the high group should consists of the upper 27 percent of the total group and the lower group of 27 percent.

In page (92) the same person has given the item response pattern as follows-

<table>
<thead>
<tr>
<th>Response</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Difficulty Discrimination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper 27%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower 27%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Freeman (1965),** indicated that the most marked and significant discrimination between the two extreme groups is obtained when item analysis is obtained based upon the highest 27% and the lowest 27% of the group.


Item screening is the rejection of some poor items and retaining some good items on the basis of the two indices of item difficulty and item discrimination. But different investigators have their own choices of screening points.

Jalota (1972),* followed 20% difficulty index and .20 discrimination index for screening items for constructing Group test of general mental ability.

Lokesh Koul (1984),** suggested that the items with the validity indices of 0.20 or more and difficulty indices of 0.40 to 0.60 to be regarded as satisfactory and items having zero or negative validity must be discarded.

Freeman (1965),*** stated that the difficulty levels can be given in terms of standard deviation of the normal curve. Thus if 84 percent of the testees pass an item, it has a rank of -1 s.d. (one s.d. below the mean) if an item passed by 16% its rank is +1 s. d. If passed by 69% or 31%, the ranks would be -0.5 s.d. or +0.5 s.d. respectively.

---

** Lokesh Koul, Methodology of Educational Research. Vika House Ghaziabad (U.P.), 1984, p.128.
Lindeman (1971),* has suggested to keep the difficulty value of .40 to .60 for screening items. In his book Lewis (1971),** also mentioned that a test maker should keep the item difficulty value between .20 to .80 and the discrimination power at .30 or greater.

As recommended by William and Lehmann (1984) *** the power of discrimination should be .20 and above. B. Shah also used to keep the discrimination indices ranged between 0.34 and 0.83 while screening the test items of omnibus test of intelligence.

Anastasi (1976),**** has given a unique suggestion that every easy item (even passed by 100% of the cases) which are discarded as non-discriminative in the usual standardized test, are very items that should be included in a mastery test. Similarly, a pretest, administered prior to a learning unit to determine whether any of the students have already acquired the skills to be taught, will yield very low percentages of passing for each item. In this case,

** Lewis, Op.cit.,
items with very low or even zero value should not be discarded, since the reveal that remains to be learned. It is apparent from these examples that the appropriate difficulty level of items depends upon the purpose of the test.

According to Lewis (1971),* the concept of the reliability of a test refers to its relative freedom from unsystematic errors of measurement, or the consistency of measurement under different conditions that might introduce error into scores. Test reliability \((r_{xx})\) is defined as the ratio of true variance to observed variance, or the proportion of the observed variance. The reliability of a test is expressed as a positive decimal number varying from \(0.00\) to \(1.00\) where \(r_{xx} = 1.00\) indicates perfect reliability or \(r_{xx} = 0.00\) the absence of reliability.

As it is found in William and Lehmanns (1984),** theoretically, reliability \((r_{xx})\) is defined as the ratio of the true score and observed score variances. Reliability tells us to what extent the observed variance is due to true score variance. The symbol \(r_{xx}\) is used for reliability, because so many of the reliability estimates are computed by the Person's product-moment correlation coefficient \((r)\).

---

William and Lehmanns (1984),* identified that the methods used to estimate reliability differ. Many different approaches can be used to estimate reliability, but the most common ones reported are listed as: Measures of stability
Measure of equivalence
Measures of equivalent and stability
Measures of internal consistency
(a) Split-half, (b) Kuder-Richardson estimates.

Ibid. (1984),** A measure of reliability of stability, often called a test-retest estimate of reliability, obtained by administering a test to a group of persons, readministering the same to the same group at a later date, and correlating the two sets of scores. The split-half method is ordinarily considered as a measure of internal consistency because the two equivalent forms are contained in a single test. In most cases the Person Product-moment correlation coefficient (r 1/2 1/2) is an estimate of the reliability to a test only half as long as the original.

According to Richard Lindeman (1971),*** to estimate what the reliability of the whole test would be, a correlation factor needs to be

applied. The appropriate formula is the Spearman-Brown formula:

\[ r_{xx} = \frac{2r^{1/2} - 1}{1 + r^{1/2}} \]

Where \( r^{1/2} \) = reliability of the half-test.

Norman E. Gronlund (1981),* The reliability of test can be estimated from a single administration to a single form of test. To split the test into halves that are most equivalent, the usual procedure is to score the even-numbered items and the odd-numbered items separately. This provides two scores for each pupil, which, when, correlated, provides a measure of internal consistency. Besides this, but another method is to compute Kuder-Richardson Formula 21. This formula can be applied to the results of any test that has been scored on the basis of the number of correct answers. Formula:

Reliability estimate (KR 21) = \( \frac{K}{K-1} \left( 1 - \frac{M(K-M)}{s^2} \right) \)

Where \( K \) = the number of items in the test
\( M \) = the mean of the test
\( s^2 \) = the standard deviation of the test scores

Anastasi (1976),** proposed that the reliability coefficient is sometimes called a coefficient of internal consistency. Once the two half-scores have been obtained for each person, they may be correlated.

---

by the usual method. However, this correlation actually gives the reliability of only a half-test. So coefficient can be estimated by means of Spearman-Brown formula.

William and Lehmanns (1984), * If items are scored dichotomously (right or wrong), one way to avoid the problems of how to split the test to use one of the Kuder-Richardson formula. The formula may be considered as representative of the average correlation obtained from all possible split-half reliability estimates. K-R 20 and K-R 21 are two formulas used extensively.

Lindeman (1971), ** After an appropriate intervening period of time, the same test is re-administered to the same group, and the correlation coefficient between the first and second administrations is computed. This coefficient the serves as the reliability estimate. It is often called a coefficient of stability.

Albert (1980), *** Thus if you compute K-R (21) and get a low reliability coefficient, you can ordinarily expect a higher one, if you compute K-R (20) and of course, a positively still higher 'r'.

---

According to William and Lehmann (1984), validity can be defined as the extent to which certain inferences can be made from test scores or other measurement. Some writers make a distinction between two kinds of criterion-related validity: Concurrent validity and predictive validity. When they collected at approximately the same time as the test data, we speak of concurrent validity, when they are gathered at a later date, we have a measure of positive validity.

As categorically mentioned by Lewis (1971)**, concurrent validation procedures are employed whenever a test is administered to people falling in various categories. If the average score varies substantially from category, then the test might be used as another, perhaps more efficient, means of classifying people into these categories.

In the opinion of Cornbach (1970),*** logically, predictive and concurrent validation are the same, and most writers apply the term predictive to both. Where one intends to emphasize that no time elapsed between measures, the study is spoken of as a concurrent validation. The designer of a new test will

suggest its validity by comparing it concurrently with an established test.

Lewis (1971),* stated if the correlation coefficient between tests X and Y is close to $+1.00$ then it can be predicted with confidence that an examinee who makes a high score on the test X will also score low on test Y. On the other hand, if the correlation between tests X and Y is close to $-1.00$, then it can be predicted that an examinee who score high on X will score low on the test Y, and an examinee who scores low on the test X will score high on test Y. (1) The closer the value of the correlation coefficient is to $+1.00$ or $-1.00$, the more accurate will the prediction be; (2) the closer the coefficient is to $.00$, the lesser accurate will the predictions be.

Norms are important that they tell us how others have performed on the test. The persons' scores can be compared with the scores of others. Norms provide us the standard to compare test performance. These norms serve as a frame of reference for interpreting raw scores indicating an examinee's standing on the test relative to scores obtained by the same age, grade, etc.

---

Freeman (1965).* Hence, to facilitate interpretation, sound psychological tests, will provide tables of age norms or grade norms or percentile ranks or decile ranks or standard scores, depending upon the instruments purpose. Other kinds of norms suitable to the test should be provided.

Ibid. (1965), the percentile method is a technique whereby scores on two or more tests, given in units that are different, may be transformed into uniform and comparable values.

Norman E. Gronlund (1981), ** A desirable feature of percentile norms is that we can interpret a pupil's performance in terms of any group in which he is a member or desire to become a member.

Lyman (1978), *** very often a single norms table is constructed to show results from several different groups. Besides the obvious economy in printing, this practice permits the comparison of a person's raw score with as many of these groups as we wish. Ibid. (1978), we find the percentile rank of an examinee or of a given raw-score value. We find a specified percentile value by finding its equivalent raw-score value.

According to Thorndike (1977),* For each raw-score the manual reports, instead of a specific percentile corresponding to that score, a range of a percentile values within which the true ability of the examinee may be presumed to lie.

William and Lehmann (1984)** Stanines are derived score with a mean of 5 and a standard deviation of 2. Only the integers 1 and 9 occur. In a normal distribution, stanines are related to other scores. The percentage of scores at each stanine are 4, 7, 12, 17, 20, 17, 12, 7, and 4 respectively.

Lyman (1978),*** chart 6.6 suggested, How to compute stanines: The purpose is to assign the stanines according to the designated percentages; . . . . The ideal percentage is shown for each stanine on the top line below.

<table>
<thead>
<tr>
<th>4%</th>
<th>7%</th>
<th>12%</th>
<th>17%</th>
<th>20%</th>
<th>17%</th>
<th>12%</th>
<th>7%</th>
<th>4%</th>
</tr>
</thead>
<tbody>
<tr>
<td>4%</td>
<td>11%</td>
<td>23%</td>
<td>40%</td>
<td>60%</td>
<td>77%</td>
<td>89%</td>
<td>96%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Cumulative Percentages.

Steps in Computing Stanines:

(1). Draw up a frequency distribution,

(2). Find the cumulative frequency up through each score value.

(3). Change these cumulative frequencies to percentages by multiplying every cf. value by $\frac{100}{N}$.

(4). Assign stanine values by approximating the ideal cumulative percentages on the bottom line above, as close as possible.

(5). Remember: each person with same raw score must receive the same stanine score—regardless of how well each value fits the ideal percentages.

Crow Crow (1979),* On the High School and college levels, an individual's intelligence status sometimes is expressed in terms of percentile scores, which indicates the place of an individual in a grade or age group on the basis of percentages of the group that score lower than he does.

William and Lehmann (1984),** DIQs are computed separately for each age group within the norm sample. These are not literal intelligence quotients. They are transformations much like the $z$ or $T$ values. Typically, these deviation IQs have a mean of 100 and a standard deviation of 15 or 16.

Ibid. (1984, p. 376), Intelligence scores using the 1972 norms are computed as deviation IQ


scores with a mean of 100 and a standard deviation of 16. In other words, they are derived scores (Dev. IQ = 16z + 100).

Until the 1960 revision, Stanford-Binet (S-B) IQs were ratio IQs. The authors of the 1960 revision decided to adopt the deviation IQ so that the standard deviation would be constant from age to age. In the case of the S-B, we have a linear standard score with a mean of 100 and a standard deviation of 16. Separately for each chronological age group, the authors have used the formula IQ = 16z + 100.

Norman E. Gronlund (1981)* stated that expressing IQs in terms of standard scores gives them the advantage over ratio IQs, of equal units and a common meaning at all age levels. They can also be readily converted to percentile ranks and other types of standard scores.

Doubtfully expressed by Freeman (1965)** that the deviation IQ furthermore, is especially useful at age levels above 16 or 18 years. For these and older persons, the use of mental age and for the ratio IQ of (MA/CA) have been regarded as inappropriate and questionable by many psychologists.

Following are the Scales given in different books for interpretation and classification of scores:

<table>
<thead>
<tr>
<th>IQ</th>
<th>DIQ</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>140 &amp; above</td>
<td>140 &amp; above</td>
<td>Very Superior</td>
</tr>
<tr>
<td>130-139</td>
<td>120-139</td>
<td>Superior</td>
</tr>
<tr>
<td>120-129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110-119</td>
<td>110-119</td>
<td>High Average</td>
</tr>
<tr>
<td>100-109</td>
<td>90-109</td>
<td>Average/Normal</td>
</tr>
<tr>
<td>90-99</td>
<td></td>
<td>Low Average</td>
</tr>
<tr>
<td>80-89</td>
<td>80-89</td>
<td>Borderline</td>
</tr>
<tr>
<td>70-79</td>
<td>70-79</td>
<td>Mentally</td>
</tr>
<tr>
<td>Below 70</td>
<td>Below 70</td>
<td>Mentally</td>
</tr>
</tbody>
</table>

** Dr.G.C.Ahuja, Test Manual p.16.
(II).

<table>
<thead>
<tr>
<th>Stanine</th>
<th>Percentile Ranks</th>
<th>Percentile Ranks</th>
<th>Descriptive Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1-4</td>
<td>5 or below</td>
<td>Very low; Very weak.</td>
</tr>
<tr>
<td>II</td>
<td>5-11</td>
<td>5-15</td>
<td>Low; Weak.</td>
</tr>
<tr>
<td>III</td>
<td>12-23</td>
<td>15-25</td>
<td>Below Average; Slightly weak.</td>
</tr>
<tr>
<td>IV</td>
<td>24-40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>41-59</td>
<td>25-75</td>
<td>Average; Satisfactory.</td>
</tr>
<tr>
<td>VI</td>
<td>60-76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>77-88</td>
<td>75-85</td>
<td>Above Average; Good.</td>
</tr>
<tr>
<td>VIII</td>
<td>89-95</td>
<td>85-95</td>
<td>High; Excellent.</td>
</tr>
<tr>
<td>IX</td>
<td>96-99</td>
<td>95 or above</td>
<td>Very high; Superior.</td>
</tr>
</tbody>
</table>