# Quadrant II – Transcript and Related Materials

**Programme: Bachelor of Arts**

**Subject: Psychology**

**Course Code: PSC 106**

**Course Title: Psychological Testing**

**Unit: II Norms In Psychological Testing**

**Module Name: Types of Norms**

**Name of the Presenter: Ms. Sweta Shyam Matonkar**

---

**Notes**

**Types of Norms**

Norms are the test performance data of a particular group of test takers (normative sample) that are designed for use as a reference when evaluating or interpreting individual test scores. Members of the normative sample will all be typical with respect to some characteristic(s) of the people for whom the particular test was designed. A test administration to this representative sample of testtakers yields a distribution (or distributions) of scores. These data constitute the norms for the test and typically are used as a reference source for evaluating and placing into context test scores obtained by individual testtakers. The data may be in the form of raw scores or converted scores. norming, refer to the process of deriving norms. The process of administering a test to a representative sample of testtakers for the purpose of establishing norms is referred to as standardization or test standardization.

a percentile is a ranking that conveys information about the relative position of a score within a distribution of scores. More formally defined, a percentile is an expression of the percentage of people whose score on a test or measure falls below a particular raw score.

1. **Percentile** A percentile is a converted score that refers to a percentage of testtakers. Percentage correct (performance in a test) refers to the distribution of raw scores—more specifically, to the number of items that

were answered correctly multiplied by 100 and divided by the total number of items. A problem with using percentiles with normally distributed scores is that real differences between raw scores may be minimized near the ends of the distribution and exaggerated in the middle of the distribution. This distortion may even be worse with highly skewed data. In the normal distribution, the highest frequency of raw scores occurs in the middle. That being the case, the differences between all those scores that cluster in the middle might be quite small, yet even the smallest differences will appear as differences in percentiles. The reverse is true at the extremes of the distributions, where differences between raw scores may be great, though we would have no way of knowing that from the relatively small differences in percentiles.

2. **Age norms** Also known as age-equivalent scores, age norms indicate the average performance of different samples of testtakers who were at various ages at the time the test was administered. If the measurement under consideration is height in inches, for example, then we know that scores (heights) for children will gradually increase at various rates as a function of age up to the middle to late teens.

3. **Grade norms** Designed to indicate the average test performance of testtakers in a given school grade, grade norms are developed by administering the test to representative samples of children over a range of consecutive grade levels (such as first through sixth grades). Like age norms, grade norms have great intuitive appeal. Children learn and develop at varying rates but in ways that are in some aspects predictable. Perhaps because of this fact, grade norms have widespread application, especially to children of elementary school age. Grade norms do not provide information as to the content or type of items that a student could or could not answer correctly. Perhaps the primary use of grade norms is as a convenient, readily understandable gauge of how one student's performance compares with that of fellow students in the same grade. One drawback of grade norms is that they are useful only with respect to years and months of schooling completed. They have little or no applicability to children who are not yet in school or to children who are out of school. Further, they are not typically designed for use with

adults who have returned to school. Both grade norms and age norms are referred to more generally as developmental norms, a term applied broadly to norms developed on the basis of any trait, ability, skill, or other characteristic that is presumed to develop, deteriorate, or otherwise be affected by chronological age, school grade, or stage of life.

4. **National norms** are derived from a normative sample that was nationally representative of the population at the time the norming study was conducted. In the fields of psychology and education, for example, national norms may be obtained by testing large numbers of people representative of different variables of interest such as age, gender, racial/ethnic background, socioeconomic strata, geographical location (such as North, East, South, West, Midwest), and different types of communities within the various parts of the country (such as rural, urban, suburban). The precise nature of the questions raised when developing national norms will depend on whom the test is designed for and what the test is designed to do. it is always a good idea to check the manual of the tests under consideration to see exactly how comparable the tests are. Two important questions that test users must raise as consumers of test-related information are “What are the differences between the tests I am considering for use in terms of their normative samples?” and “How comparable are these normative samples to the sample of testtakers with whom I will be using the test?”

5. **National anchor norms** Even the most casual survey of catalogues from various test publishers will reveal that, with respect to almost any human characteristic or ability, there exist many different tests purporting to measure the characteristic or ability. Dozens of tests, for example, purport to measure reading. Suppose we select a reading test designed for use in grades 3 to 6, which, for the purposes of this hypothetical example, we call the Best Reading Test (BRT). Suppose further that we want to compare findings obtained on another national reading test designed for use with grades 3 to 6, the hypothetical XYZ Reading Test, with the BRT. An equivalency table for scores on the two tests, or national anchor norms, could provide the tool for such a comparison. Just as an anchor provides some stability to a vessel, so national anchor norms provide some stability to test scores by anchoring them to other test

scores. The method by which such equivalency tables or national anchor norms are established typically begins with the computation of percentile norms for each of the tests to be compared. Using the equipercentile method, the equivalency of scores on different tests is calculated with reference to corresponding percentile scores. Thus, if the 96th percentile corresponds to a score of 69 on the BRT and if the 96th percentile corresponds to a score of 14 on the XYZ, then we can say that a BRT score of 69 is equivalent to an XYZ score of 14. We should note that the national anchor norms for our hypothetical BRT and XYZ tests must have been obtained on the same sample—each member of the sample took both tests, and the equivalency tables were then calculated on the basis of these data. technical considerations entail that it would be a mistake to treat these equivalencies as precise equalities.

6. **Local norms** Typically developed by test users themselves, local norms provide normative information with respect to the local population's performance on some test. A local company personnel director might find some nationally standardized test useful in making selection decisions but might deem the norms published in the test manual to be far afield of local job applicants' score distributions. Individual high schools may wish to develop their own school norms (local norms) for student scores on an examination that is administered statewide. A school guidance center may find that locally derived norms for a particular test—say, a survey of personal values— are more useful in counseling students than the national norms printed in the manual. Some test users use abbreviated forms of existing tests, which requires new norms. Some test users substitute one subtest for another within a larger test, thus creating the need for new norms. There are many different scenarios that would lead the prudent test user to develop local norms.

7. **Subgroup norms** A normative sample can be segmented by any of the criteria initially used in selecting subjects for the sample. What results from such segmentation are more narrowly defined subgroup norms. Thus, for example, suppose criteria used in selecting children for inclusion in the XYZ Reading Test normative sample were age, educational level, socioeconomic level, geographic region, community type, and

handedness (whether the child was right-handed or left-handed). The test manual or a supplement to it might report normative information by each of these subgroups. A community school board member might find the regional norms to be most useful, whereas a psychologist doing exploratory research in the area of brain lateralization and reading scores might find the handedness norms most useful.

Norms provide a context for interpreting the meaning of a test score. Another type of aid in providing a context for interpretation is termed a fixed reference group scoring system. Here, the distribution of scores obtained on the test from one group of testtakers—referred to as the fixed reference group—is used as the basis for the calculation of test scores for future administrations of the test. Perhaps the test most familiar to college students that has historically exemplified the use of a fixed reference group scoring system is the SAT. This test was first administered in 1926. Its norms were then based on the mean and standard deviation of the people who took the test at the time. With passing years, more colleges became members of the College Board, the sponsoring organization for the test. It soon became evident that SAT scores tended to vary somewhat as a function of the time of year the test was administered. In an effort to ensure perpetual comparability and continuity of scores, a fixed reference group scoring system was put into place in 1941. The distribution of scores from the 11,000 people who took the SAT in 1941 was immortalized as a standard to be used in the conversion of raw scores on future administrations of the test.5 A new fixed reference group, which consisted of the more than 2 million testtakers who completed the SAT in 1990, began to be used in 1995. A score of 500 on the SAT corresponds to the mean obtained by the 1990 sample, a score of 400 corresponds to a score that is 1 standard deviation below the 1990 mean, and so forth. As an example, suppose John took the SAT in 1995 and answered 50 items correctly on a particular scale. And let's say Mary took the test in 2008 and, just like John, answered 50 items correctly. Although John and Mary may have achieved the same raw score, they would not necessarily achieve the same scaled score. If, for example, the 2008 version of the test was judged to be somewhat easier than the1995 version, then scaled scores for the 2008 testtakers would be calibrated downward. This would be done so as to make scores earned in 2008 comparable to scores earned in 1995. Test items common to each new version of the SAT and each previous version of it are employed in a procedure (termed anchoring) that permits the conversion of raw scores on the new version of the test into fixed reference group scores. Like

other fixed reference group scores, including Graduate Record Examination scores, SAT scores are most typically interpreted by local decision-making bodies with respect to local norms. Thus, for example, college admissions officers usually rely on their own independently collected norms to make selection decisions. They will typically compare applicants' SAT scores to the SAT scores of students in their school who completed or failed to complete their program. Of course, admissions decisions are seldom made on the basis of the SAT (or any other single test) alone. Various criteria are typically evaluated in admissions decisions.

**References**


1. Cohen, J. R. & Swerdlik, M. E. (2018). Psychological Testing and Assessment: An Introduction to Tests and Measurement. (9th ed.). New Delhi: McGraw-Hill Education.
2. Gregory, R. J. (2017). Psychological Testing: History, Principles, and Applications (7th Ed.).New Delhi: Pearson (India) Pvt. Ltd.