

## Norms

Norms refers to information regarding the group performance of a particular reference on a particular measure for which a person can be compared to.

- Sir Francis Galton at the first time developed the logic for norm based testing in the 18<sup>th</sup> century.
- Norms mean standardized score.
- A norm represents a typical level of performance for a particular group.
- A raw score on any Psychological test alone is meaningless unless we have additional interpretive data.
- So the score on psychological test are most commonly interpreted by reference to norms that represent the test performance of the standardized sample.
- Norms always represent the best performance.
- Norms are empirically established by determining what persons in a representative group actually do on a test.
- In order to ascertain more precisely the individual's exact position with reference to the standardized sample, the raw score is converted into some relative measure.

These derived scores serve two purposes.

1. They indicate the individual's relative standing in the normative sample and facilitate evaluation of performance.
2. They provide comparable measures that permit a direct comparison of the individuals' performance on different tests.

## Types of Norms

Fundamentally, derived scores are expressed in one of two major ways

- (1) Development Norms and
- (2) Within group Norms.

**(1) Developmental Norms:-** These type of norms generally indicate the normal developmental path the individual has progressed. They are very helpful for descriptive purpose but they are not compatible to precise statistical treatment.

The types of development norms are

- (a) Mental Age Norms,
- (b) Grade Equivalent Norms and
- (c) Ordinal Scale Norms.

**Mental Age:** The term “mental age” was widely popularized through the various translations and adaptations of the Binet-Simon scales, although Binet himself had employed the more neutral term “mental Level”.

- In age scales such as the Binet and its revisions (prior to 1986), items were grouped into year levels.
- For example, those items passed by the majority of 7-year olds in the standardization sample were placed in the 7-year level, and so forth.
- A child’s score on the test would then correspond to the highest year level that he or she could successfully complete. In actual practice, the individual’s score is often below their mental age and passed some above it. For this reason, it was customary to compute the basal age, that is, the highest age at and below which all tests were passed. Partial credits, in months, were then added to this basal age for all tests passed at higher year levels. Mental age norms have also been employed with tests that are not divided into year levels. In such case, the child’s raw score is first determined.
- The mean raw scores obtained by the children in each year group within the standardization sample constitute the age norms for such a test.
- The mean raw score of the 8-year old children, for example, would represent the 8-year old raw score then her or his mental age on the test is 8 years.
- All raw scores on such a test can be transformed in a similar manner by reference to the age norms.

**Grade Equivalents:** Scores on educational achievement tests are often interpreted in terms of grade equivalents.

- Grade norms are found by computing the mean raw score obtained by children in each grade.
- Thus, if the average number of problems solved correctly on an arithmetic test by the fourth graders in the **standardization** sample is 23, then a raw score of 23 corresponds to grade equivalents of 4.

- Intermediate grade equivalents, representing fractions of a grade, are usually found by interpolation, although they can also be obtained directly by testing children at different times within the school years.
- For example, 4.0 refer to average performance at the beginning of the fourth grade.
- Grade norms are also subject to misinterpretation unless the test user keeps firmly in mind the manner in which they were derived.

**Ordinal Scales:** Ordinal scales are designed to identify the stage reached by the child in the development of specific behavior functions.

- Although scores may be reported in terms of approximate age levels, such scores are secondary to qualitative **description** of the child's characteristics behavior.
- The ordinarily of such scales refers to the uniform progression of development through successive stages in so far as these scales typically provide information about what the child is actually able to do(e.g. climbs stairs without assistance; recognizes identity in quantity of liquid when poured into differently shaped containers), they share important features with the domain-referenced tests.

**(2) Within Group Norms:-**

- Such types of norms help in comparing the individual's performance with the most nearly comparable standardized group's performance.
- The individual's performance is evaluated in terms of the performance of the most nearly comparable standardization group.
- Within group norms have a uniform and clearly defined quantitative meaning and can be appropriately employed in most types of statistical analyses.

**a. Percentiles (P(n) and PR):-** Percentile scores represent the percentage of persons in the standardized sample who fall below a given raw score.

- They indicate an individual's relative position in the standardized sample. In case of percentiles, the counting begins from the bottom so lower the percentile, poorer the standing / rank.

**b. Standard Score:** - Standard score express the individual's distance from the Mean in terms of the standard deviation of the distribution.

- They are obtained by linear or nonlinear transformation of the original raw scores.
- T Scores and Z scores are known as standard scores.

**c. Age Norms :-** To establish age norms, the Mean of raw scores obtained by all in the same age group within a standardized sample is taken.

- So Mean raw score of 12 year old students would represent the 12 years norm.

**d. Grade Norms :-** Grade norms are found by computing the Mean row score obtained by students in particular grade.

**The normative sample:** any norm, however, expressed, is restricted to the particular normative population from which it was derived.

**Specific Norms:** to standardize tests on more narrowly defined populations, so chosen as to suit the specific purpose of each test.

**Fixed Reference Group:**

- One type of non-normative scale utilizes a fixed reference group in order to ensure comparability and continuity of scores, without providing normative evaluation of performance.
- With such a scale, normative interpretation requires reference to independently collected norms from a suitable population. Local or other specific norms are for used for this purpose.

### **Domain-Referenced Test Interpretation / Criterion-Referenced Test**

- Several alternative terms are in common use, such as content-, domain-, and objective-referenced.
- Domain-referenced, or criterion-referenced, test interpretation is the concept that an examinee's scores on a test are interpreted with reference to the particular cognitive

ability being assessed rather than in comparison with the performance of a population of individuals (norm-referenced testing).

- Example: when examining reading ability, domain-referenced interpretation is exemplified when a neuropsychologist interprets an examinee's performance in terms of reading mastery or how the current reading level compares to a previous level of reading. This interpretation also assists in predicting the examinee's performance on a criterion.

<b>Dimension</b>	<b>Criterion-Referenced Tests</b>	<b>Norm-Referenced Tests</b>
<b>Purpose</b>	<p>To determine whether each student has achieved specific skills or concepts.</p> <p>To find out how much students know before instruction begins and after it has finished.</p>	<p>To rank each student with respect to the achievement of others in broad areas of knowledge.</p> <p>To discriminate between high and low achievers.</p>
<b>Content</b>	<p>Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts.</p> <p>Each skill is expressed as an instructional objective.</p>	<p>Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts.</p>
<b>Item Characteristics</b>	<p>Each skill is tested by at least four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing.</p> <p>The items which test any given skill are parallel in difficulty.</p>	<p>Each skill is usually tested by less than four items.</p> <p>Items vary in difficulty.</p> <p>Items are selected that discriminate between high and low achievers.</p>
<b>Score Interpretation</b>	<p>Each individual is compared with a preset standard for acceptable achievement. The performance of other examinees is irrelevant.</p> <p>A student's score is usually expressed as a percentage.</p> <p>Student achievement is reported for individual skills.</p>	<p>Each individual is compared with other examinees and assigned a score--usually expressed as a percentile, a grade equivalent score, or a stanine.</p> <p>Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills.</p>

The following is adapted from: Popham, J. W. (1975). *Educational evaluation*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.