# Quadrant II – Transcript and Related Materials

**Programme: Bachelor of Arts (Third Year)**

**Subject: Psychology**

**Course Code: PSC106**

**Course Title: Psychological Testing**

**Unit: 03 (Reliability and Validity)**

**Module Name: Concept of reliability, sources of error variance**

**Name of the Presenter: Michelle Fernandes (Ph.D)**

## Notes

*Reliability* is a synonym for *dependability* or *consistency.* In psychometrics *reliability* refers to consistency in measurement. A test may be reliable in one context and unreliable in another. There are different types and degrees of reliability. A **reliability coefficient** is an index of reliability, a proportion that indicates the ratio between the true score variance on a test and the total variance. In this chapter, we explore different kinds of reliability coefficients, including those for measuring test-retest reliability, alternate-forms reliability, split-half reliability, and inter-scorer reliability.

According to the classical test theory, a score on an ability test is presumed to reflect not only the testtaker's true score on the ability being measured but also error. In its broadest sense, error refers to the component of the observed test score that does not have to do with the testtaker's ability. If we use *X* to represent an observed score, *T* to represent a true score, and *E* to represent error, then the fact that an observed score equals the true score plus error may be expressed as follows:

$X = T + E$

A statistic useful in describing sources of test score variability is the **variance** (σ2)—the standard deviation squared. This statistic is useful because it can be broken into components. Variance from true differences is **true variance,** and variance from irrelevant, random sources is **error variance.** If σ2 represents the total variance, the true variance, and the error variance, then the relationship of the variances can be expressed as

$$\sigma^2 = \sigma^2_{th} + \sigma^2_e$$

In this equation, the total variance in an observed distribution of test scores ($\sigma^2$) equals

the sum of the true variance ($\sigma^2_{th}$) plus the error variance ($\sigma^2_e$). The term **reliability** refers to the proportion of the total variance attributed to true variance. The greater the proportion of the total variance attributed to true variance, the more reliable the test. Because true differences are assumed to be stable, they are presumed to yield consistent scores on repeated administrations of the same test as well as on equivalent forms of tests. Because error variance may increase or decrease a test score by varying amounts, consistency of the test score—and thus the reliability—can be affected.

In general, the term **measurement error** refers to, collectively, all of the factors associated with the process of measuring some variable, other than the variable being measured.

Measurement error, can be categorized as being either *systematic* or *random.* **Random error** is a source of error in measuring a targeted variable caused by unpredictable fluctuations and inconsistencies of other variables in the measurement process. Sometimes referred to as "noise," this source of error fluctuates from one testing situation to another with no discernible pattern that would systematically raise or lower scores.

In contrast to random error, **systematic error** refers to a source of error in measuring a variable that is typically constant or proportionate to what is presumed to be the true value of the variable being measured. For example, a 12-inch ruler may be found to be, in actuality, a tenth of one inch longer than 12 inches. All of the 12-inch measurements previously taken with that ruler were systematically off by one-tenth of an inch; that is, anything measured to be exactly 12 inches with that ruler was, in reality, 12 and one-tenth inches. In this example, it is the measuring instrument itself that has been found to be a source of systematic error.

***Sources of Error Variance***

Sources of error variance include test construction, administration, scoring, and/or interpretation.

**Test construction** One source of variance during test construction is **item sampling** or **content sampling,** terms that refer to variation among items within a test as well as to variation among items between tests.

**Test administration** Sources of error variance that occur during test administration may influence the testtaker's attention or motivation. The testtaker's reactions to those influences are the source of one kind of error variance. Examples of untoward influences during administration of a test include factors related to the *test environment:* room temperature, level of lighting, and amount of ventilation and noise, for instance.

Other potential sources of error variance during test administration are *testtaker variables.* Pressing emotional problems, physical discomfort, lack of sleep, and the effects of drugs or medication can all be sources of error variance. Formal learning experiences, casual life experiences, therapy, illness, and changes in mood or mental state are other potential sources of testtaker-related error variance.

*Examiner-related variables* are potential sources of error variance. The examiner's physical appearance and demeanor—even the presence or absence of an examiner—are some factors for consideration here. Some examiners in some testing situations might knowingly or unwittingly depart from the procedure prescribed for a particular test.

**Test scoring and interpretation:** In many tests, the advent of computer scoring and a growing reliance on objective, computer-scorable items have virtually eliminated error variance caused by scorer differences.

Scorers and scoring systems are potential sources of error variance. A test may employ objective-type items amenable to computer scoring of well-documented reliability.

**Other sources of error** Surveys and polls are two tools of assessment commonly used by researchers who study public opinion. In the political arena, for example, researchers trying to predict who will win an election may sample opinions from representative voters and then draw conclusions based on their data. However, in the "fine print" of those conclusions is usually a disclaimer that the conclusions may be off by plus or minus a certain percent. This fine print is a reference to the margin of error the researchers estimate to exist in their study. The error in such research may be a result of sampling error—the extent to which the population of voters in the study actually was representative of voters in the election. The researchers may not have gotten it right with respect to demographics, political party affiliation, or other factors related to the population of voters. Alternatively, the researchers may have gotten such factors right but simply did not include enough people in their sample to draw the conclusions that they did. This brings us to another type of error, called methodological error. So, for example, the interviewers may not have been trained properly, the wording in the questionnaire

may have been ambiguous, or the items may have somehow been biased to favor one or another of the candidates.

**References:**

Cohen, R. J., & Swerdlik, M. E. (2018). Psychological testing and assessment: An introduction to tests and measurement. (9th ed.). New Delhi: McGraw-Hill Education.