The course code is PSC106, the title of the unit is reliability and validity, and the title of the module that we will be doing is, concept of reliability and sources of error variance.

I am Dr. Michelle Fernandes from Carmel College of Arts, Science and Commerce for Women.

In this module we'll be talking about the concept of reliability. We will also look at the reliability coefficient. What are measurement errors and the various sources of measurement error.

By the end of this module, you should be able to define what is reliability. You should also be able to comprehend the concept of a reliability coefficient. You should be able to distinguish, understand and recognize various measurement errors and distinguish the various sources of measurement errors that can take place in a psychological testing.

So, let's start, let's define the concept of reliability.

Now in everyday conversation, reliability is a synonym for dependability or consistency. So, when we talk about reliability, we talk about whether a test is dependable, whether it is consistent. What do you mean by consistent, consistent means that every time you measure using the test, you should get similar results on the same person?

Now, broadly speaking, in the language of psychometrics, reliability refers to consistency in measurement. Now in every day conversation reliability generally means something positive, but in psychometric sense it really only refers to something that is consistent, not necessarily whether it is good or bad, but something that is consistent

Reliability is not an all-or-none matter. A test may be reliable in one context and unreliable in another context.

Now a very common definition, a very well-known definition was proposed by an Anastasi in 1968, and she defines reliability as the consistency of scores obtained by the same individual when re-examined with test on different occasions or with different sets of equivalent items or under other variable examining conditions.

Now, in determining reliability, it is assumed that the test is measuring a relatively stable characteristic. Unlike instability, unreliability is the result of measurement errors produced by temporary internal states, such as low motivation, indisposition or external conditions such as distractions or an uncomfortable testing environment.

Now, what is the reliability? Coefficient a reliability coefficient is an indexed of reliability, a proportion that indicates the ratio between true test score variance on a test and the total variance. All types of reliability can be expressed in terms of the correlation coefficient.

Now this correlation coefficient is expressed as the Greek letter rho. Now, the letter is also known when correlation coefficient is also referred to when the population is unknown. Rho is the sample correlation coefficient.

Correlation coefficient is expressed as the degree of correspondence or relationship between two sets of scores. A correlation has a value of +1.00 to – 1.00. A zero correlation indicates complete absence of a relationship. Correlation coefficients may be computed in various ways, depending on the nature of the data, the most common is the Pearson product-moment correlation coefficient.

Now when we talk about reliability, there is also another concept called as the classical test theory. Now, according to the classical test theory, the score on an ability test is presumed to reflect not only the test takers true score on the ability being measured, but also the error. In its broadest sense error refers to the components of the observed test scope that does not have to do with the test takers ability.

Now, if, for example X is used to represent an observe score, T is used to represent the true scope and E to represent the error, then an observed score may be expressed as follows. X= T+E. The lesser the random error, the more the raw scores represent the true score.

Now, if you look at this, observed score or the score that you obtain from a psychological test will be a combination of an error score and the true scope. So, this observed score will have both the error score as well as the true score.

Now reliability theories show that the variance of obtained score is simply the sum of variance of the true scores plus the variance of error measurement. This equation suggests that test scores vary as a result of two factors. Variability in true scores and variability due to errors in measurement.

Now what is a true score? A true score is the replicable feature of the concept being measured. It is the part of the observed score that would recur across different measurement occasions in the absence of an error.

Error measurements are composed of both random error and systematic error. It represents the discrepancies between scores obtained on tests and the corresponding true score.

A statistic useful to describing sources of test score variability is the variance – 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 the variance square represents the total variance, the true variance and the error variance, then the relationship of the variance can be expressed as $\sigma^2 = \sigma^2_{th} + \sigma^2_e$

Now, 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 the true variance, the more reliable the test. Because the 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 the tests. Because error variances may increase or decrease the test score by varying amounts, consistency of the test score and thus reliability, can be affected.

Now, 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, much like any error in general, can be categorized as being either systematic or random. A 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.

Examples of random error that could conceivably affect test scores range from unanticipated events happening in the immediate vicinity of test environment such as lightning strikes or a spontaneous occupancy of the university rally, etc to unanticipated physical events happening within the test taker, such as sudden or unexpected surge in the test takers, blood sugar or blood pressure.

In contrast, to random error, systematic errors refer to the 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 actually 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. Once a systematic error becomes known, it becomes predictable—as well as fixable.

Now what are the various sources of error variance?

The sources of error variance would include test construction, administration, scoring and interpretation and other sources of error.

Now let's look at test construction as a source of error variance. 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.

Consider two or more test designed to measure a specific skill, personality attribute, or a body of knowledge. Differences are sure to be found in the way the items were worded and in the exact context sampled. Each of us has probably walked into an achievement test setting thinking “I hope they don't ask this question” or “I hope they don’t ask that question.” If the only questions on the examination were the ones we hoped would be asked, we might achieve a higher score on that test.

Test administration: Sources of error variance that occur during test administration may influence the test taker’s attention or motivation. The test taker’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 so on and so forth. It could also refer to test taker variables like for example, a person who is not feeling well.

Some examiner related variables, where the examiner could influence the test taker. Examiner related variables are potential sources of error variance. The examiner's physical appearance. The examiner might knowingly or unknowingly depart from the procedure prescribed for a particular test. Examiners may provide clues to the examiners and the level of professionalism exhibited by the examiners.

Another source of error variance could be test scoring and interpretation Now, with the advent of computer scoring and the growing reliance on objective, computer based scorable items have virtually eliminated error variance caused by scorer differences. Individually administered intelligence test, some types of personality tests, tests of creativity, various behavioral measures, and countless other tests still require hand scoring by a trained examiner, and this can cause error variance.

If subjectivity is involved in scoring, then the scorer may be the source of error variance. The element of subjectivity in scoring may be much greater in the administration of

certain non-objective-type personality tests, tests of creativity, and certain academic tests. Subjectivity in scoring can even enter into behavioral assessment.

The other sources of error variance would be certain types of assessment situations that can lend themselves to particular varieties of systematic and non-systematic error. For example, consider assessing the extent of agreement between partners regarding the quality and quantity of physical and psychological abuse in their relationship.

Now potential sources of non-systematic error in such an assessment situation include forgetting, failing to notice abusive behavior, and misunderstanding instructions regarding reporting. A number of studies have suggested that underreporting or overreporting of abuse also may contribute to systematic error.

Females, for example, may underreport abuse because of fear, shame and so on and so forth. Males may underreport abuse because of embarrassment and social desirability factors and overreport abuse if they are attempting to justify the report.

Now, in this module we spoke about the concept of reliability and we also spoke about the various sources of error that can take place in a psychological test.

For this module you can refer to the following books.

I hope you enjoyed this module.

Thank you.