

Session 10

Characteristics of a Valid Assessment Tool

A. Introduction



Image 1: Southeast Asia rank: PH 2nd to worst in Grade 5 students' reading, math (Source: Dela Peña, 2023: Online)

The lives and futures of young people are profoundly impacted by the assessment and testing they undergo. The prospects and possibilities available to students are affected by decisions made inside and by their institutions, but the outcomes of their performance on standardized exams and examinations hold an even higher degree of significance. When the

outcomes of tests and examinations are used to pass judgments on teachers and schools, they also affect how students are taught at such institutions. Because of their significance, it is necessary that the findings of summative assessment accurately reflect and have some impact on the education received in schools.

When evaluating the value of a test, validity is among the most crucial factors to consider. Validity refers to the extent to which the test assesses the constructs it purports to (Navarro, Santos & Corpuz, 2019). The items in a valid test should strongly relate to the evaluated topic. This means that many certification and licensing exam questions will be particularly job- and industry-specific. A valid test would accurately measure the skills and knowledge relevant to the position being filled. This invalidates the use of the test results for their intended function.

Session 9 Conclusion

The multiple-choice test is versatile because it encompasses almost all cognitive learning skills. The action of the learning objective determines the test stem and the answers following the procedure: (1) begin with the expected Answer, (2) Draft the stem or question, and (3) assign worthy distracters.

There are a few different ways to ensure the validity of the test given to students to assess learning and teaching. This lecture will cover only the validity of the objective type test, particularly the multiple-choice test.

B. Session Objectives

After studying this module, you are expected to:

1. Determine the various aspects of validating an assessment instrument and
2. Construct valid multiple-choice test items across the first five cognitive skills: Remembering, Understanding, Applying, Analyzing, and Evaluating.

C. Session Content

1. Types of Test Validity

The test's validity is the degree to which a test measures what it claims to measure. According to Kubiszyn and Borich (2007), validity can also refer to the appropriateness, accuracy, meaningfulness, and utility of a teacher's specific decisions based on the test results. The objective of the validation process is to acquire information regarding the qualities of the examination as a whole. The procedure of gathering and examining evidence to back up the meaningfulness and usefulness of the test is known as evidence-gathering and analysis.

The evaluator gathers necessary evidence when checking the validity of the test items or the whole test paper. In this case, there are at least four types of tests to ensure the validity of the assessment instrument: content, criterion-related, construct, and face.

Content Validity. The ability of a test to represent the various components of a given construct is what's meant by its "content validity," which is an evaluation criterion. It is crucial that the content of the survey, test, or any measuring method that you use cover the relevant and necessary parts of the subject it wants to assess. Only then will you be able to get reliable results from your efforts?

The test paper should contain test items to measure the pupils' learning across the subject matters and competencies covered during the examination period. If some components of the measurement are missing, or if you included irrelevant aspects, the validity is in jeopardy.

The test constructor should ensure that the test items are congruent to the test coverage. For example, if the pupils' learning of the lesson content being assessed is about the layers of the earth and their theoretical properties, test items should cover topics about the crust, mantle, and core. Those items are invalid if there are items regarding the earth's atmosphere. Not unless the earth's atmosphere is also another content to be assessed.

Construct Validity. The concept of construct validity refers to how accurately a test measures the topic for which it was developed. It's essential for determining a method's overall credibility.

Researching intangible concepts like intelligence, self-confidence, and happiness necessitates carefully evaluating construct validity. To avoid the probability of bias, it is necessary to use numerous observable or measurable indicators to assess these notions.

Construct and content validity are highly related. Nonetheless, construct validity ensures that the learning concept for a specific content is contextually relevant to the test item. For instance, if the learning concept to be assessed is "to describe how the mantle's asthenosphere causes the crust to move", then the test item should be derived directly from the statement of the learning concept. However, if the item asks about the composition of the mantle, then it has a construct validity issue even if the same content is covered.

Criterion-related Validity. Responses are compared to future performance or those gathered from other, more reliable surveys to determine criterion validity.

There are two types of criterion validity: predictive and contemporaneous. When discussing test instruments, predictive validity is the degree to which they can accurately predict outcomes. Predictive validity can be demonstrated by an entrance exam accurately identifying those most likely to thrive in advanced study. When there is agreement between two assessments or when a new measure compares favorably to an established valid measure, this is evidence of concurrent validity. Concurrent validity can be found, for instance, when a further study is compared to an existing validated measure using the same sample of respondents or when a novel instrument is used to collect data and the results are compared to expert opinion. Concurrent validity can be defined as a strong relationship

between the new survey and the existing standard. Concurrent validity is helpful when a new, purportedly superior measure is developed.

Test instruments in schools can be validated through criterion-related validity, correlating the results of one test with the results of another of the same students. For example, the scores of pupils in their summative Mathematical problem-solving tests can be correlated with the standardized problem-solving test. In this way, the tested skill in the summative test is compared with the results of a similar or related item in the standardized test. If the scores have a significantly positive correlation, the test is valid.

Face Validity. A test's "face validity" is how well it seems to measure the idea or quality it's supposed to. You could judge how well it fits the subject or how important it is. It is based on the personal opinion of the assessor. It doesn't need statistical study to be figured out. Face validity is a quick and easy way to check how true a measure seems. It does not, however, consistently show that the test is valid (Kubiszyn & Borich, 2007).

Face validity is much simpler when it comes to test papers to be administered to students. It is mainly on the physical presentation of the test, e.g., layout, design, margin, font style, and size, as well as the types of tests, grammar, semantics, and the choice of words. These characteristics should be according to the interests and academic levels of the test takers.

3. Constructing Valid Test Items

Test items are the tasks and/or questions used to create tests and examinations in test development. The most effective methods for developing test items draw from the science of test design and the art of testing. On the one hand, the entire process of creating a high-quality product includes a sizeable amount of experimentation that we might classify as scientific. Still, writing itself requires a great deal of original thought. At the end of the procedure, a novel, potent, and helpful product emerges: a test instrument that repurposes words, symbols, and other resources from a course of study for various pedagogical ends. The designer's imagination and ingenuity are just as crucial as their knowledge: the item's structure, proportion, and balance are all factors.

This lecture also focuses on ensuring that the test items are valid according to the student's cognitive skills to be assessed. Therefore, the test constructor must know the attributes and distinctions of Bloom's taxonomy of cognitive learning.

The following examples show the learning outcomes according to the first five cognitive skills: remembering, understanding, applying, analyzing, and evaluating. We will focus on the first three steps in constructing multiple-choice test items: (1) beginning with the expected or correct Answer, (2) crafting the stem or question, and (3) assigning the distracters.

1. Remembering

Consider that the learning objective is “to describe the tympanic membrane or eardrum.” Let’s say that the expected answer is “eardrum”. A remembering test item may simply be “What is the film-like sheath at the end of the ear canal that is also known as the tympanic membrane?”

The content is valid because the scope of learning is consistently limited only to the tympanic membrane. As regards construct, it measures the recall skill of the student by recalling that the eardrum is also called the tympanum.

Observe the following examples of how the learning outcomes derive the valid test items.

2. Understanding

Learning Outcome: To describe the structure and location of the malleus, incus, and stapes

Expected Answer: malleus

Stem: Which bone is attached to the tympanic membrane that receives the sound waves?

3. Applying

Learning Outcome: To determine how the ossicles work in the ear

Expected Answer: to amplify sound into the inner ear

Stem: What do the auditory ossicles do in the hearing sensory process of the ear?

4. *Analyzing*

Learning Outcome: To infer how the Eustachian tube maintains the balance in the auditory system

Expected Answer: Maintains the body's equilibrium

Stem: Why do you think the Eustachian tube balances the pressure on both ears?

5. *Evaluating*

Learning Outcome: To discuss the proper care of the ear

Expected Answer: Exposure to very loud noises

Stem: Which of the following should be avoided to prevent damage to the ear?

Note: *Creating* cannot be assessed using any objective-type test.

D. Conclusion

The test item is valid when the content is based on the learning outcome being assessed and is proven to measure the learning skill intended in its construction. Other than the content and construct validity, the entire test's physical features and language proficiency make face validity equally essential.

E. References

Kubiszyn, T. & Borich, G. (2007). *Educational testing and measurement: Classroom application and practice. 8th edition*. N.J., U.S.A.: John Wiley & Sons, Inc.

Navarro, R.L., Santos, R.G., & Corpuz, B.B. (2019). *Assessment in learning 1. 4th edition*. Quezon City, Philippines: LORIMAR Publishing

Dela Peña, K. (2023). *Southeast Asia rank: PH 2nd to worst in Grade 5 students' reading, math* [Online Image] [Accessed on October 15, 2023] <https://newsinfo.inquirer.net/1721616/southeast-asia-rank-ph-2nd-to-worst-in-grade-5-students-reading-math-skills>