UOT 81
https://doi.org/10.59849/2311-8482.2024.2.74

ROYA ZEYNALOVA

BASIC CONCEPTS OF LANGUAGE ASSESSMENT

There are five main concepts in determining language assessment: Language assessment is a measure of the proficiency a language user has in any given language. It could be a first or second language. Tests are one form of language assessment. The assessment may include listening, speaking, reading, writing, an integration of two or more of these skills, or other constructs of language ability. Equal weight may be placed on knowledge (understanding how the language works theoretically) and proficiency (ability to use the language practically), or greater weight may be given to one aspect or the other.

Keywords: *language, concept, assessment, listening, speaking*

Introduction

Language assessment is a measure of the proficiency a language user has in any given language. It could be a first or second language. Tests are one form of language assessment and there are many others. They fall into two categories: summative and formative. According to Brown there are five fundamental criteria for testing a test: practically, reliability, validity, authenticity, and washback. Let's point out those are:

Main content

Practicality The first characteristic of an effective test is practicality. Practicality relates to the considerations of cost of a test, time allotment, test administration, human resource, test construction, and test scoring (1). A good test should be relatively low in cost. It should be affordable by the students or test-takers.. A test which is prepared in a power point display for the whole class can be cheaper compared with the use of paper, but it may not be practical because it is difficult for the students who need to think longer or faster, or for the students who want to look back again at the previous items. Find a test which is low in cost, but does not sacrifice the quality of the test. A test that is prohibitively expensive is impractical. A test of language proficiency that takes a student five hours to complete is impractical-it consumes more time than necessary to accomplish its objective. A test that requires individual one-on-one proctoring is impractical for a group of several hundred test-takers and only a handful of examiners.(4) A test that takes a few minutes for a student to take and several hours for an examiner to evaluate is impractical for most classroom situations. A test that can be scored only by computer is impractical if the test takes place a thousand miles away from the nearest computer. The value and quality of a test sometimes hinge on such nitty-gritty, practical considerations.

Reliability - The second characteristic of a good test is reliability. Reliability means consistency, i.e. consistency in relation to students or test-takers, raters or scorers, test administration, and the test itself.(5) There are several factors which affect assessment reliability. To get reliable scores from the test-takers, we need to be sure that the test-takers are in good physical and mental conditions when taking the test. A test-taker who is unfit, fatigue, or in bad mood at the time of taking the test, may not be able to concentrate, and therefore cannot show his/her best or real performance. In other words, the result of his/her test may not be reliable. The test-takers who are not familiar with the procedure of doing the test will not be able to reach optimal performance in the test either. This, in turn, makes the result of the test unreliable. Unreliable test results may also be shown when in a group of test-takers some of them are familiar with the test procedure that they can do the test faster and more easily, while others who are not familiar with the test procedure do the test in confusion and uncertainty.(2) The raters or scorers of a test should possess reliability. They should. A reliable test is consistent and dependable. If you give the same test to the same student or matched students on two different occasions, the test should yield similar results. The issue of reliability of a test may best be addressed by considering a number of factors that may contribute to the unreliability of a test.

Validity -By far the most complex criterion ofan effective test and arguably the most important principle is validity, "the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment". A valid test of reading ability actually measures reading ability not 20/20 vision, nor previous knowledge in a subject, nor some other variable of questionable relevance. To measure writing ability, one might ask students to write as many words as they can in 15 minutes, then simply count the words for the final score. Such a test would be easy to administer (practical), and the scoring quite dependable (reliable).(3) But it would not constitute a valid test of writing ability without some consideration of comprehensibility, rhetorical discourse elements, and the organization of ideas, among other factors. How is the validity of a test established? There is no final, absolute measure of validity, but several different kinds of evidence may be invoked in support. In some cases, it may be appropriate to examine the extent to which a test calls for performance that matches that of the course or unit of study being tested. In other cases, we may be concerned with how well a test determines whether or not students have reached an established set of goals or level of competence. Statistical correlation with other related but independent measures is another widely accepted form of evidence. Other concerns about a test's validity may focus on the consequences beyond measuring the criteria themselves of a test, or even on the test-taker's perception of validity.(6)

Content-related evidence -If a test actually samples the subject matter about which conclusions are to be drawn, and if it requires the test-taker to perform the behavior that is being measured, it can claim content-related evidence of validity, often popularly referred to as content validity. You can usually identify content-related evidence observationally if you can clearly define the achievement that you are measuring. A test of tennis competency that asks someone to run a 100-yard dash obviously lacks content validity. If you are trying to assess a person's ability to speak a second language in a conversational setting, asking the learner to answer", paper and pencil multiple-choice questions requiring grammatical judgments does not achieve content validity.(1) A test that requires the learner actually to speak within some sort of authentic context does. And if a course has perhaps ten objectives but only two are covered in a test, then content validity suffers.

There are a few cases of highly specialized and sophisticated testing instruments that may have questionable content-related evidence of validity. It is possible to contend,for example, that standard language proficiency tests, with their contextreduced, academically oriented language and limited stretches of discourse, lack content validity since they do not require the full spectrum of communicative performance on the part of the learner. There is good reasoning behind such criticism; nevertheless, what such proficiency tests lack in content-related evidence they may gain in other forms of evidence, not to mention practicality and reliability. Another way of understanding content validity is to consider the difference between direct and indirect testing. Direct testing involves the test-taker in actually perfoming the target task. In an indirect test, learners are not performing the task itself but rather a task that is related in some way. For example, if you intend to test learners' oral production of syllable stress and your test task is to have learners mark (with written accent marks) stressed syllables in a list of written words, you could, with a stretch of logic, argue that you are indirectly testing their oral production. A direct test of syllable production would have to require that students actually produce target words orally. The most feasible rule of thumb for achieving content validity in classroom assessment is to test performance directly. Consider, for example, a listening and speaking class that is doing a unit on greetings and exchanges that includes discourse for asking for personal information (name, address, hobbies, etc.) with some form-focus on the verb to be, personal pronouns, and question formation. The test on that unit should include all of the above discourse and grammatical elements and involve students in the actual performance of listening and speaking.(5) What all the above examples suggest is that content is not the only type of evidence to support the validity of a test,but classroom teachers have neither the time nor the budget to subject quizzes, midterms, and final exams to the extensive scrutiny of a full construct validation. Therefore, it is critical that teachers hold content-related evidence in high esteem in the process of defending the validity of classroom tests.

Criterion-related evidence -A second form of evidence of the validity of a test may be found in what is called criterion-related evidence, also referred to as criterion-related validity, or the extent to which the "criterion" of the test has actually been reached. In such tests, specified classroom objectives are measured, and implied predetermined levels of performance are expected to be reached.In the case of teacher-made classroom assessments, criterion-related evidence is best demonstrated through a comparison of results of an assessment with results of some other measure of the same criterion. For example, in a course unit whose objective is for students to be able to orally produce voiced and voiceless stops in all possible phonetic environments, the results of one teacher's unit test might be compared with an independent assessment-possibly a commercially produced test in a textbook of the same phonemic proficiency. A classroom test designed to assess mastery of a point of grammar in communicative use will have criterion validity if test scores are corroborated either by observed subsequent behavior or by other communicative measures of the grammar point in question. Criterion-related evidence usually falls into one of two categories: concurrent and predictive validity. A test has concurrent validity if its results are supported by other concurrent perfonnance beyond the assessment itself. For example, the validity of a high score on the final exam of a foreign language course will be substantiated by actual proficiency in the language. The predictive validity of an assessment becomes important in the case of placement tests, admissions assessment batteries, language aptitude tests, and the like. The assessment criterion in such cases is not to measure concurrent ability but to assess (and predict) a test-taker's likelihood of future success.(5)

Construct-related evidence -A third kind of evidence that can support validity, but one that does not playas large a role for classroom teachers, is ,construct-related validity, commonly referred to as construct validity. A construct is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions. Constructs mayor may not be directly or empirically measured-their verification often requires inferential data. "Proficiency" and "communicative competence" are linguistic constructs; "self-esteem" and "motivation" are psychological constructs. Virtually every issue in language learning and teaching involves theoretical constructs. In the field of assessment, construct validity asks, "Does this test actually tap into the theoretical construct as it has been defined?" Tests are, in a manner of speaking, operational definitions of constructs in that they operationalize the entity that is being measured.For most of the tests that you administer as a classroom teacher, a formal construct validation procedure may seem a daunting prospect. You will be tempted, perhaps, to run a quick content check and be satisfied with the test's validity. But don't let the concept of construct validity scare you. An informal construct validation of the use of virtually every classroom test is both essential and feasible. Imagine, for example, that you have been given a procedure for conducting an oral interview. The scoring analysis for the interview includes several factors in the final score. So if you were asked to conduct an oral proficiency interview that evaluated only pronunciation and grammar, you could be justifiably suspicious about the construct validity of that test. Likewise, let's suppose you have created a simple written vocabulary quiz, covering the content of a recent unit, that asks students to correctly define a set of words. (6)

Construct validity is a major issue in validating large-scale standardized tests of proficiency. Because such tests must, for economic reasons, adhere to the principle of practicality, and because they must sample a limited number of domains of language, they may not be able to contain all the content of a particular field or skill. The TOEFL, for example, has until recently not attempted to sampie oral production, yet oral production is obviously an important part of academic success in a university course of study. The TOEFL's omission of oral production content, however, is osten· sibly justified by research that has shown positive correlations between oral production and the behaviors (listening, reading, grammaticality detection, and writing) actually sampled on the TOEFL.Because of the crucial need to offer a fmancially affordable proficiency test and the high cost of administering and scoring oral -production tests, the omission of oral content from the TOEFL has been justified as an economic necessity.(1)

Consequential validity -As well as the above three widely accepted forms of evidence that may be introduced to support the validity ofan assessment, two other categories may be of some interest and utility in your own quest for validating classroom tests. Messick, Gronlund, McNamara and Brindley, among others, underscore the potential importance of the consequences of using an assessment. Consequential validity encompasses all the consequences of a test, including such considerations as its accuracy in measuring intended criteria, its impact on the preparation of test-takers, its effect on the learner, and the (intended and unintended) social consequences of a test's interpretation and use.(3) As high-stakes assessment has gained ground in the last two decades, one aspect of consequential validity has drawn special attention: the effect of test preparation courses and manuals on performance. McNamara cautions against test results that may reflect socioeconomic conditions such as opportunities for coaching.

Face validity -An important facet of consequential validity is the extent to which "students view the assessment as fair, relevant, and useful for improving learning" , or what is popularly known as face validity. Face validity refers to the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers.

Authenticity

The next principle is authenticity, a base that covered the design a form of test, including the features, appropriate language, and the implication of the test. The tendency of this principle may be students feasible recognized the language related to fact or not just perception. It might present in to the following ways(4)

- Use natural language;
- The items prior contextualized;
- Meaningful topics (relevant, interesting)
- The items organize in thematic way (through a story line or episode)
- Represent real-world task.

Washback

Washback can be defined as the effect of test or assessment on teaching, learning, learner, or government and society. Washback can be positive or negative. For example, since there is a writing test in the national examination, teachers who were previously reluctant to teach writing, then they teach writing. Knowing that the test is always challenging to the students, then the students are motivated to learn and make better preparation for the test. These are examples of positive washback. However, when teachers know that the national examination always uses multiple choice test items, then in the teaching and learning activities the teachers drill their students on how to do multiple choice test, forgetting teaching students the process of learning, this is an example of negative washback. Or, knowing multiple choice examination, students are busy preparing the effective strategy for cheating.(6) This is the worst negative washback. Washback occurs more in classroom assessment when information could 'washes back' to students and it useful to identify strengths and weaknesses. It's challenging for the teacher to achieve that washback. Many teacher, because inattention or fatigue instead just give a letter grade or score. The way to enhance washback by comment generously and specifically on test performance, such as: give complement for the strengths, constructive criticism for weaknesses, emphasized certain elements that might improve their test performance and so forth.

Conclusion

There are five principles of language assessment; they are practicality, reliability, validity, authenticity, and washback.An effective test is practical. This means that it is not excessively expensive, stays within appropriate time constraints, is relatively easy to administer, and has a scoring/evaluation procedure that is specific and time-efficient. Another major of principle of language testing is washback, generally refers to the influence of testing on teaching and learning. authenticity, a base that covered the design a form of test, including the features, appropriate language, and the implication of the test. Reliability - A reliable test is consistent and dependable. By far the most complex criterion ofan effective test and arguably the most important principle is validity, "the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of the assessment".

REFERENCES

1. H.Douglas Brown. (2004), Language Assesment: Principles and Classroom Practices, White Plains, NY: Pearson Education.
2.Diane Larsen-Freeman (2000),Techniques and principles in language teaching. Oxford University Press.
3. Alderson, J. Charles. (2001). Language testing and assessment (Part 1). Language Teaching, 34, 213-236.
4. Alderson, J. Charles. (2002). Language testing and assessment (Part 2). Language Teaching, 35, 79-113.
5. Bailey, Kathleen M. (1998). Learning about language assessment: Dilemmas, decisions, and directions. Cambridge, MA: Heinle & Heinle.
6. H.Douglas Brown.(2007),Principles of language learning and teaching. White Plains, NY: Pearson Education.

Naxçıvan Dövlət Universiteti, müəllim
E-mail: roya.q90@gmail.com

Röya Zeynalova

DİLİN QİYMƏTLƏNMƏSİNİN ƏSAS KONSEPSİYASI

Dil qiymətləndirməsinin müəyyən edilməsində beş əsas anlayış var. Dilin qiymətləndirilməsi hər hansı bir dildə istifadəçinin dil biliyinin ölçüsüdür. Bu, birinci və ya ikinci dil ola bilər. Testlər dilin qiymətləndirilməsinin bir formasıdır. Qiymətləndirmə dinləmə, danışma, oxuma, yazma, bu bacarıqlardan iki və ya daha çoxunun inteqrasiyası və ya dil qabiliyyətinin digər strukturlarını əhatə edə bilər. Biliyə (dilin nəzəri cəhətdən necə işlədiyini anlamaq) və biliyə (dili praktiki olaraq istifadə etmək bacarığına) bərabər çəki verilə bilər və ya bu və ya digər aspektə daha çox əhəmiyyət verilə bilər.

Açar sözlər: *dil, anlayış, qiymətləndirmə, dinləmə, danışma*

Роя Зейналова

ОСНОВНЫЕ КОНЦЕПЦИИ ОЦЕНКИ ЯЗЫКА

Существует пять основных концепций определения языковой оценки: Языковая оценка – это мера владения языком, которым владеет пользователь любого конкретного языка. Это может быть первый или второй язык. Тесты являются одной из форм языковой оценки. Оценка может включать в себя аудирование, говорение, чтение, письмо, интеграцию двух или более из этих навыков или другие составляющие языковых способностей. Одинаковый вес может быть придан знанию (пониманию того, как язык работает теоретически) и владению языком (способности использовать язык на практике), или больший вес может быть отдан тому или иному аспекту.

Ключевые слова: *язык, концепция, оценка, слушание, говорение*

AMEA-nın müxbir üzvü Əbülfəz Quliyev tərəfindən təqdim edilmişdir

İlk daxilolma tarixi: 08.05.2024
Son daxilolma tarixi:29.06.2024