ФИЛОЛОГИЧЕСКИЕ НАУКИ

UDK 81'33

Sultana Begzada M.Rafiq
Literature and Human Sciences Faculty

Faryab University
Afghanistan, Maymana

THE QUALITIES AND USEFULNESS OF LANGUAGE TESTS

The most important quality of a test is its usefulness. This may seem so obvious that it need not be stated. But what makes a test useful? How do we know of a test will be useful before we use it? Or if it has been useful after we have used it? Stating the question of usefulness this way implies that simply using a test does not make it useful.

We believe that test usefulness provides a kind of metric by which we can evaluate not only the tests that we develop and use, but also all aspects of test development and use. We thus regard a model of test usefulness as the essential basis for quality control throughout the entire test development process. We would further argue that all test development and use should be informed by a model of test usefulness. In this article we propose a model of test usefulness that includes six test qualities- reliability, construct validity, authenticity, instructiveness, impact, and practicality. We also propose three principles that we believe are the basis for operationalizing our model of usefulness in the development and use of language test.

Key words: *qualities, reliability, construct validity, authenticity, instructiveness, impact, practicality, principles, tests.*

TEST USEFULNESS

The traditional approach to describing test qualities has been to discuss these as more or less independent characteristics, emphasizing the need to maximize them all. This has led some language testers to what we see as the extreme and untenable position that maximizing one quality leads to the virtual loss of others. Language testers have been told that the qualities of reliability are essentially in conflict [4; 8].

Our notion of usefulness can be expressed as in figure 1.

Usefulness = reliability + Construct validity + Authenticity + Instructiveness + Impact + Practicality

Principle 1	It is the overall usefulness of the test that is to be maximized, rather than the individual qualities that affect usefulness.
Principle 2	The individual test qualities cannot be evaluated independently, but must be evaluated in terms of their combined effect on the overall usefulness of the test.
Principle3	Test usefulness and the appropriate balance among the different qualities cannot be prescribed in general, but must be determined for each specific testing situation.

Figure 1. Usefulness

The development and use of language tests is provided by the three principles that follow.

These principles reflect our belief that, in order to be useful, any given language test must be developed with a specific purpose, a particular group of test takers and a specific language use domain in mind.

Test Qualities

The main difference between tests and other components of an instructional program, in our view, is in their purpose. While the primary purpose of other components is to promote learning, the primary purpose of tests is to measure. Tests can serve pedagogical purposes, to be sure, but this is not their *primary* function. Four of the qualities that we will discuss with respect to tests are shared by other components of a learning program. Thus, we can consider the authenticity of a particular language sample that may be used for instruction, the instructiveness of a particular learning task, the impact of a given learning activity, or the practicality of a particular teaching approach for a given situation. Two of the qualities-reliability and validity – are, however, critical for tests, and are sometimes referred to as essential measurement qualities. This is because these are the qualities that provide the major justification for using test scores- numbers – as a basis for making inferences or decisions [3].

The Consideration in Designing a Language Test

The most important consideration in designing a language test is its usefulness, and this can be defined in terms of six test qualities: reliability, validity, authenticity, interactiveness, impact, and practicality. These six test qualities all contribute to test usefulness, so that they cannot be evaluated independently of each other. Furthermore, the relative importance of these different qualities will vary from on testing situation on another, so that test usefulness can only be evaluated for specific testing situations. Similarly, the appropriate balance of these qualities cannot be prescribed in the abstract, but can only be determined for a given test. The most important consideration to keep in mind is not to ignore any one quality at the expense of others. Rather, we need to strive to achieve an appropriate balance, given the purpose of the test, the characteristics of the TLU domain and the test takers, and the way we have defined the construct to be measured [7].

1. ***Reliability*** can be defined as consistency of measurement; inconsistency is variation in test scores that is due to factors other than the construct we want to measure. Of the many factors that can affect test performance, the ones over which we have some control are the characteristics of test tasks. Thus, in designing and developing language tests, we try to minimize variations in the test task characteristics that are not motivated by the way in which we have defined the construct and TLU tasks. In addition to attempting to minimize such unmotivated variations through design, we need to estimate their effects on test scores, to determine how successful we have been in minimizing them as sources of inconsistency of measurement.

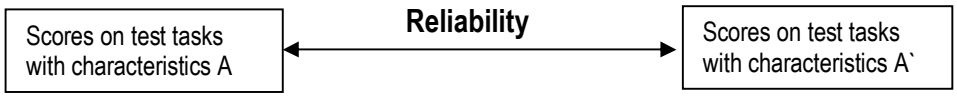

Figure 2. Reliability

In this figure, the double-headed arrow is used to indicate a correspondence between two sets of task characteristics (A and A`) which differ only in incidental ways**.**

2. *Construct validity* pertains to the meaningfulness and appropriateness of the interpretations that we make on the basis of test scores. The validity of these interpretations cannot simply be asserted, but must be demonstrated. Test validation is the ongoing process of demonstrating that a particular interpretation. The justification that we need to provide is evidence of construct validity, or evidence that the test score reflects the area of language ability we want to measure, and very little else [2].

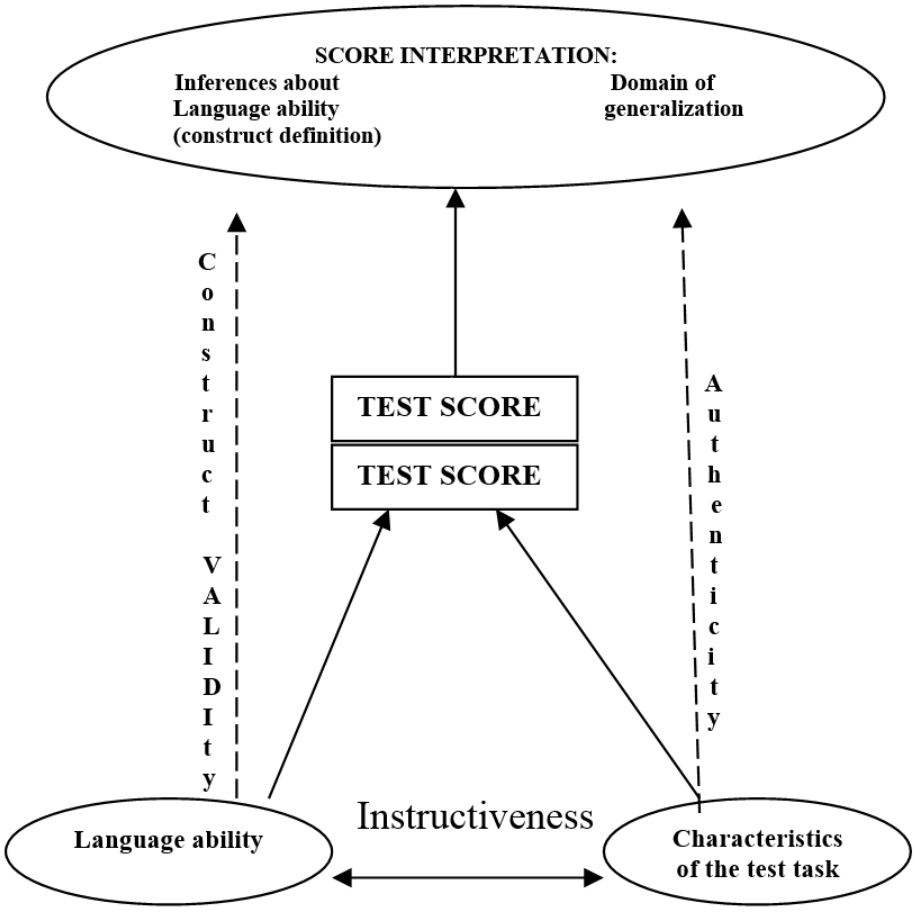

Figure 3. Construct validity of score interpretations

3. *Authenticity* is an important quality for language tests for two reasons:

a. It provides a link between test performance and the TLU tasks and domain to which we want to generalize, and

b. The way test takers perceive the relative authenticity of test tasks can, potentially, facilitate their test performance.

We define authenticity as the degree of correspondence of the characteristics of a given language test task to the characteristics of a TLU task. The relationship is shown in figure 4 [9].

Figure 4. Authenticity

4. ***Instructiveness*** is an important test quality because it pertains to the degree to which the constructs we want to assess are critically involved in accomplishing the test task. Instructiveness is also important because it is at the heart of many current views of language teaching and language learning. instructiveness is a function of the extent and type of involvement of the test taker's language ability (language knowledge plus metacognitive strategies), topical knowledge, and affective schemata in accomplishing a test task.

Both authenticity and instructiveness are relative, so that we speak of 'relatively more' or 'relatively less' authentic and interactive, rather than 'authentic' and 'inauthentic', or 'interactive' and 'non-interactive'. Furthermore, we cannot determine the relative authenticity or instructiveness of a test task just by looking at it; we must also consider the characteristics of the test takers, the TLU domain, and the test task.

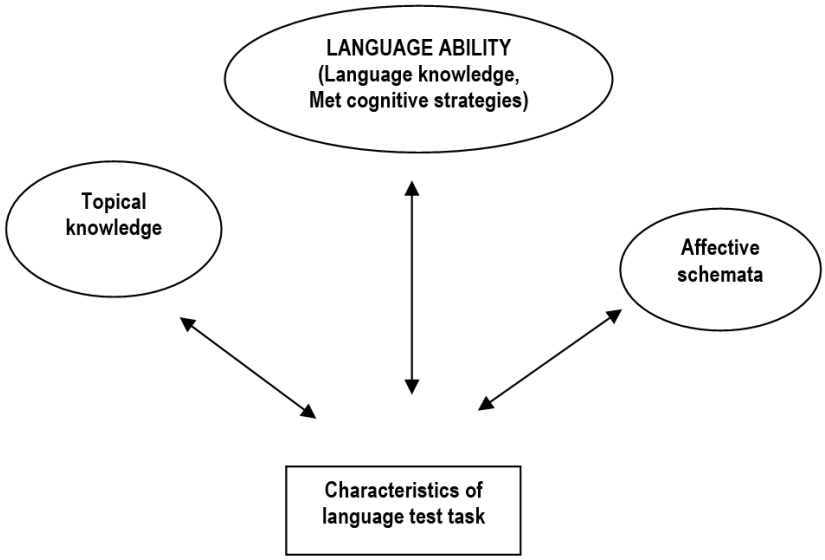

Figure 5. Instructiveness

5. ***Impact*** can be defined broadly in terms of the various ways in which test use affects society, and education system, and the individuals within these. As test developers and test users we must always consider the societal, educational, and individual value systems that inform our test use. We also need to think through carefully what might happen as a result of our using a test for a particular purpose. Impact operates at two levels: a macro level, in terms of the societal or educational system in general, and a micro level, in terms of the individuals who are affected by the particular test use. The notion of washback, or backwash, which has been of considerable concern in language testing, can be viewed in terms of various aspects of impact [1; 6].

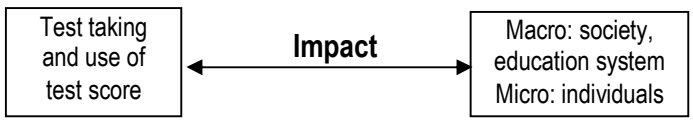

Figure 6. Impact

6. Unlike the other five qualities, which pertain to the uses that are made of test scores, ***practicality*** pertains to the ways in which the test will be implemented in a given situation, or whether the test will be used at all. We can define practicality as the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities. A practical test is one whose design, development, and use do not require more resources than are available. Several *types of resources* can be identified: human resources, material resources, and time. The specific resources required will vary from one situation to another, as will the resources that are available. Thus, practicality can only be determined for a specific testing situation, and it makes little sense to say that a given test or test task is more or less practical than another, in general. Considerations of practicality are likely to affect our decisions at every stage in the process of test development and use, and these may lead us to reconsider and perhaps revise some of our earlier specifications. In designing and developing a test, we try to achieve the optimum balance among the qualities of reliability, construct validity, authenticity, instructiveness, and impact, for our particular testing situation. In addition, we must determine the resources required to achieve this balance, in relationship to the resources that are available.

Conclusion

We believe the approach to defining test usefulness developed here makes a contribution to the field of language testing for two reasons. First, it provides a principled basis for considering the relative importance of all the qualities that contribute to usefulness, and enables us to consider how these interact with each other. Second, it ties the notion of usefulness to the specific testing situation. That is, it links considerations of reliability, validity, authenticity, instructiveness, impact, and practicality to the specific purpose of the test, to a specific TLU domain, to specific groups of test takers and test users, and to specific local conditions with respect to the availability and allocation of resources. This approach to test usefulness thus makes two requirements of test developers. First, we must consider these qualities with respect to specific tests, and not solely in terms of abstract theories and statistical formulate. Second, we must consider these qualities from the very beginning of the test planning and development process, and rather than relying solely on *ex post facto* analyses.

REFERENCES

1. Alderson J.C., Wall D. 'Does washback exist?' // Applied Linguistics. – 1993. – № 14. – P. 115–129.
2. Bachman's L.F. *Fundamental Considerations in Language Testing.* – Oxford University Press, 1990. – P. 285–289.
3. Bachman F.L., Plamer. *Language Testing in Practice.* – Oxford University Press, 1996. – P. 17–19.
4. Heaton G.B. *Writing English Language Tests.* – Second edition. – London: Longman, 1988.
5. Hughes A. *Testing for Language Teachers.* – Cambridge University Press, 1989.
6. Messick S. 'Validity' in R. L. Linn (ed.). *Educational Measurement.* – Third edition. – New York: American Council on Education and Macmillan, 1989. – P. 13–103.
7. Swain M. 'Large-scale communicative language testing: A case study' // Y. P. Lee, A. C. Y. Fok, R. Lord, G. Low (eds.). *New Directions in Language Testing.* – Oxford: Pergamon Press, 1985. – P. 35–46.
8. Underhill N. 'The great reliability/validity trade-off: Problems in assessing the productive skills' // J. B. Heaton (ed.). *Language Testing.* – London: Modern English Publications, 1982.
9. Widdowson H.G. *Learning Purpose and Language Use.* – Oxford University Press, 1983.

Султана Бегзада М.Рафик
Факультет литературы и гуманитарных наук

Фарьябский Университет
Афганистан, г. Маймана

КАЧЕСТВО И ПОЛЕЗНОСТЬ ЯЗЫКОВЫХ ТЕСТОВ

Самым важным качеством теста является его полезность. Это настолько очевидно, что нет необходимости говорить об этом. Но что делает тест полезным? Как мы узнаем, что тест будет полезен, прежде чем использовать его? Или он был полезен после того, как мы его использовали? Постановка вопроса о полезности таким образом подразумевает, что простое использование теста не делает его полезным.

Мы считаем, что полезность тестов обеспечивает своего рода показатель, с помощью которого мы можем оценивать не только тесты, которые мы разрабатываем и используем, но также и все аспекты разработки и использования тестов. Таким образом, мы рассматриваем модель полезности тестов как существенную основу для контроля качества на протяжении всего процесса разработки тестов. Мы также утверждаем, что все разработки и использование тестов должны основываться на модели полезности тестов. В этой статье мы предлагаем модель полезности теста, которая включает шесть качеств теста – надежность, валидность конструкции, аутентичность, обучаемость, воздействие и практичность. Мы также предлагаем три принципа, которые, по нашему мнению, являются основой для практического применения нашей модели полезности при разработке и использовании языкового теста.

Ключевые слова: *качество, надежность, конструктивная валидность, аутентичность, обучаемость, воздействие, практичность, принципы, тесты.*