Research Article

		Linguistics
Authenticity in Language Testing		**Keywords:** authenticity, subjectiveness, relativity.

Yujing Zheng		**Chongqing Jiaotong University, School of Foreign Languages, 66# Xuefu Road, Nan An District, Chongqing, China 400074**
Arburim Iseni		**State University of Tetova, Macedonia**

Abstract

Since the communicative language testing was proposed in 1980s, authenticity, as a distinctive feature of it, has been a heated topic in modern language testing. This paper conducts a brief analysis of the concept of authenticity in language testing, and investigates such two characteristics of it as subjectiveness and relativity. It also comments on the authenticity in scoring. The authors hold that authenticity is a relative concept and should be considered to achieve appropriate balance with the other qualities (i.e. reliability, construct validity, interactiveness, impact, practicality) of language testing.

1. Introduction

Authenticity as a critical quality of language tests has been debated among language testing researchers over several decades (Alderson & Wall, 2000). Nevertheless, there is disagreement about the nature of authenticity and about the level of authenticity that can be achieved in language tests (Weir, 1990). Generally speaking, the theories include the real-life (RL) approach (Bachman, 1990) and correspondence approach (Bachman & Palmer, 1996). Thus, in this paper, these theories will be discussed in details and some comments will be made on them, through which, two main characteristics about authenticity will be summarized, and in the last part, the authenticity in scoring will also be discussed as a complementary concept to the theories about authenticity.

2. Theories on Authenticity in Language Testing

2.1 Real-life (RA) Approach

Different definitions of authenticity in language testing reflect different views on the nature of it. The real-life approach to defining authenticity essentially considers the extent to which test performance replicates some specified non-test language performance (Bachman, 1990, p301).

Its primary concerns are: 1) the appearance and test takers' perception of the test and how this affect the test takers' performance; 2) to what extent can the test predict the non-test performance of the test takers.

There are three interrelated tenets characterizing the RL approach: 1) a view of language ability, or proficiency, as pragmatic aspiration; 2) the reference to real-life performance as a criterion; 3) the belief that face validity, content relevance, and predictive utility are sufficient bases to justify test use. (Bachman, 1990)

The RL approach has been under attack from several aspects. First, the definition of real-life performance is not clear. In foreign language testing, what does the "real life" refer to, the real life of the native speakers or the real life of the foreign language users (test takers)?

Second, in most cases, real-life performance cannot be duplicated directly to be test performance because once a real-life task is sampled as test task, some of its features will be changed more or less. This difficulty of replicating real-life performance in language tests has also been recognized by proponents of the RL approach. For example, Davies (1977) described authenticity as a continuum between "direct" and "indirect" by using the phrases "duplicate as closely as possible…real-life situations". However, he didn't state clearly how close was close for the duplication of the real-life situation. Third, another problem with RL approach is its failure to distinguish language ability from language performance. Defining proficiency as the actual performance of it limits test interpretation and use to the prediction of future situations.

2.2 Correspondence Approach

Bachman & Palmer (1996) developed the concept of authenticity as "the degree of correspondence between the characteristics of a given language test task to the features of a TLU (Target Language Use) task" (p.23). They held that the correspondence was the primary concern of authenticity. The relationship can be shown in Figure 1
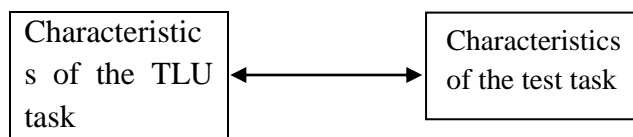
Figure 1: Authenticity (Bachman & Palmer, 1996)

The correspondence approach was built based on the framework of task characteristics. The framework defines the characteristics of test tasks as the characteristics of the setting, test rubrics, the input, the expected response and relationship between input and response.

The framework makes it possible in practice to design an authentic test task by identifying the critical features of the TLU task and developing correspondent test tasks or selecting sample tasks that have these critical features. The framework can also serves as a checking list to assess authenticity of test tasks by comparing the characteristics of it with those of the TLU tasks.

Unlike the unclear definition of "real-life" performance in the RL approach, the definition of the "TLU domains" is conceptually clear in the correspondence approach, avoiding any misunderstanding about it.

In the correspondence approach, Bachman & Palmer (1996) defined "target language use domain as a set of specific language use tasks that the test taker is likely to encounter outside of the test itself" (p.44). It was also stated that according to the purpose of language use, TLU domains consist of real-life domains, in which language was used for the purpose of

communication, and language instruction domain in which language was used for the purpose of teaching and learning.

This expansion of definition of TLU domains makes it more scientific for it takes into account some tests for the purpose of language teaching and learning and highlights the importance of the relationship between language testing and language teaching and learning.

However, there are still some drawbacks with the correspondence approach.

Firstly, although the "TLU domains" are defined conceptually, it is still difficult for the test designers to carry out this approach in practice. For example, what is the target-language use domain for a general proficiency test – all possible situations? And even if we have a more specific target, such as academic language use, the target language will vary according to the field of study, the type of study and so forth. It is often not clear which tasks we should choose.

Secondly, even if we decide which tasks are important, it is not possible to accurately replicate target-language use tasks, simply because the testing situation is different from the real-world communicative situation-it is a test (Douglas, 1999). We need to ensure, therefore, that the test task replicates the most critical aspects of the target-language use task, and the obvious way to do that is to compare the competencies they both require. And that brings us to the difficulty of determining which competencies particular tasks require.

3. Characteristics of Authenticity

From the theories discussed above, such two characteristics of authenticity can be summarized as subjectiveness and relativity.

3.1 Subjectiveness

Authenticity is subjective in that it depends on the subjective decisions on the part of the test developers when designing the test, as well as the subjective understanding on the part of the test takers when taking the test.

As mentioned above, according to Bachman & Palmer (1996), authenticity concerns the correspondence of the characteristics of the test tasks to the characteristics of the TLU domain tasks. In order to make a test authentic, the test developers must make such decisions as identifying the distinctive features in the TLU domain and then selecting the sample test tasks that have these critical features. These decisions are made based on the test developers' judgments, which are subjective more or less. Bachman & Palmer (1996) also realized this problem, they stated, "in either designing new tests or analyzing existing tests, our estimates of authenticity are only guesses" (p.29). However, they tried to minimize the subjectiveness of authenticity by developing a framework of language task characteristics, which indeed provide some bases for the

test developers to make decisions, but the problem of subjectiveness is still not avoided because the framework itself is also somewhat subjective.

The subjectiveness of authenticity can also be attributed to the test takers.

In his approach, Bachman (1990) defines authenticity as a function of the interaction between the test taker and the test task. Therefore, the test takers' perceptions of authenticity are very important to evaluate the authenticity of the test. But "different test takers may have different perceptions of the test task" (Bachman & Palmer, 1996, p.24).

In other words, the perceptions of the test task are also subjective judgments on the part of the test takers. Similarly, test takers' perceptions of the test task may also be different from those of the test developers, which is to say that even for one test task, the test developers may think it is authentic, but the test takers may regard it as not authentic. Maybe that is the reason why Bachman & Palmer (1996) stated,

> *We can do our best to design test tasks that we believe will be authentic for a given group of test takers, but we need to realize that different test takers may process the same test task in different ways, often in ways we may not anticipate.* (p.29)

3.2 Relativity

Besides subjectiveness, the concept of authenticity also has such implication as "relativity", "so that we speak of 'relative more' or 'relative less' authentic, rather than 'authentic' and 'inauthentic'" (Bachman & Palmer, 1996, p.28)

We could say that a test task is high in authentic because of the correspondence between characteristics of the TLU domain and characteristics of the test task, but never could we say that those two kinds of characteristics are totally correspondent, because as we have argued in the previous parts, any TLU domain task, when it is sampled as the test task, no matter how similar or relevant the two kinds of tasks are, some of its features must be changed to fit for the specific testing situation. So the best the test developers can do is to make a test task as authentic as possible, but not totally authentic.

The second reason why we consider authenticity as relative, it is one of the six qualities (i.e. reliability, conduct validity, instructiveness, impact, practicality) of the test usefulness (Bachman & Palmer, 1996) which makes it necessary for the test developers to take all the six qualities into consideration and try to achieve the appropriate balance among them. For this reason, the test developers or evaluators cannot pursue only authenticity despite of the other qualities, which means that the authenticity of the test is interrelated with the other qualities and the realization of it is relative.

4. Authenticity in Scoring

Authenticity does not only refer to authenticity in developing the test task and the test taker's interaction with the test task, it also include authenticity in scoring. For example, McNamara (1996) proposed that the scoring criteria should also be authentic, by which he means the scoring of the test task should also be correspondent to the assessment of the fulfillment of the tasks in the real word (qtd in Weigle, 1996). For example, if the task is to write a persuasive essay, the rating criteria for it should be how successful the writer fulfillS the task, that is, how successful he is to persuade the reader, the scoring criteria that focus on discrete aspects of language-use of vocabulary, organization, and so on-are not adequate, because such scoring criteria would make a test focus on the language used in the test task, not on the fulfillment of the task. In other words, using the discrete scoring criteria, the raters are interested more in the linguistic aspects of the writing than in whether they feel persuaded, which fails to follow the authenticity in rating.

Authentic rating is an illustration of the speech act theory, which emphasizes the effect of the behavior (i.e. writing the composition as a fulfillment of the test task). It is true that in the real-world language use, people judge the fulfillment of a communicative task by assessing whether the effect of the task (behavior) is successful or not. For example, in the real-world language use, if one takes such a writing task as writing a letter of apology to another person, we can judge his success in fulfilling the task by whether he is forgiven or not. But it seems difficult for test task to use this criterion as scoring criterion. Let's also take the task of writing an apology letter for example. As a writing test task, we definitely cannot score the examinees in this writing task just according to the subjective judgment of raters forgiving the examinees or not, because such kind of scoring is totally unreliable and unfair to the test takers. So it is definitely unwise to pursue the so-called authentic scoring at the expense of the reliability of the scoring.

5. Conclusion

With the development of the communicative language testing, authenticity have become the center of many test researchers' concerns, however, there is disagreement about the theoretical framework of authenticity. Thus, this paper analyzed the RL approach and Correspondence approach to authenticity, and summarized two main characteristics about authenticity based on the theoretical analysis. In the last part, it is discussed authenticity in scoring to make some complementary analysis to the framework of authenticity.

References

Alderson, J. C., Clapham, C., Wall, D. (2000). *Language Test Construction and Evaluation* Beijing: Foreign Language Teaching and Research Press.

Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman L.F.& Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Davies, A. (1977). The Construction of Language Tests (Chapter 3) In J. P. B. Allen & A. Davies (Eds.), Testing and Experimental Methods (pp. 38-104). Oxford: Oxford University Press.

Douglas, D. (1999). *Assessing Languages for Specific Purposes.* London: Cambridge University Press.

Weir, C. J. (1990). *Communicative Language Testing*. London: Prentice Hall.

Weigle, S.A. (1996). *Assessing Writing*. London: Cambridge University Press.