How to Design a Relatively Authentic Language Test?

Jixiu Xing
Heilongjiang Polytechnic, Heilongjiang, China.

Abstract
It is impossible to design an authentic language test, but how to design a relatively authentic one? Following the introduction, the writer depicts the notion of authenticity, the importance of authenticity, and introduces another key research area in language testing - construct validity, as well as the relationship between construct validity and authenticity. And then talks about the characteristics and dimensions of authenticity and tries to put forth a way to design a relatively authentic language test via four examples. At last, the article discusses the problems that should be paid attention to when designing and implementing an authentic test and the relativity of authenticity. It argues that, despite the positive attitude of test designers and researchers, only some of the problems can be solved and many questions remain unanswered. The article concludes that much more research is needed for test researchers to design a relatively authentic language test.

Keywords
Authentic test, construct validity, authenticity, relativity

1. Introduction
The purpose of language testing is to elicit responses to language use from the learners so that we can draw conclusions that predict the learners’ ability to use language in contexts other than the individual test situation (Alfallajand Al-Ahdal, 2017). As a result, a test similar to using language in real situations was developed in the past decades. This is a so-called authentic language test. Lewkowicz (2000) claims that the concept of authenticity has been debated both in the fields of general education and applied linguistics, and it is this concept that has reverberated around the world, with an increasing number of language testing researchers and developers following in its footsteps. However, opinions vary widely as to whether it is possible to design a truly authentic language test. After deliberating upon the question, it is not difficult to conclude that an authentic test is nonexistent. This can be demonstrated by the following study.
2. What is authenticity

To understand what an authentic test is, authenticity should be known first. Perceptions about the notion of authenticity vary from linguist to linguist. In 1961, Carroll put forward the criteria of authenticity, namely, facile performance, normal communication situation, and total communicative effect of an utterance (Bachman, 1990). Breen (1985) declares that authenticity is a concept related to texts, assignments, and communal circumstances of a classroom.

In the early 1990s, Bachman integrated the ideas of Breen and Widdowson and proposed that it was necessary to divide authenticity into two types: situational authenticity (matching the characteristics of the test tasks to the TLU tasks) and interactional authenticity (the involvement of test-takers in the test task) (Bachman, 1991). However, in 1996, Bachman and Palmer separated the notion of authenticity from interactivity and affirmed that authenticity was the degree of correspondence between characteristics of both a language test task and a target language use (TLU) task. Thus, interactional authenticity was replaced by interactivity because of a recognition that all real-life assignments are authentic in the situation. In 2000, Douglas still divided authenticity into two aspects and asserted that it was necessary to introduce them in language tests for specific purposes.

To determine the authenticity of test tasks (the degree of correspondence between language tests and TLU tasks), Bachman and Palmer (1996) suggest a framework on task characteristics that offers a systematic way to match tasks according to the setting, test input, test rubrics, test result that the tasks are expected to bring about, as well as the relationship between the input and the response. It is important because it not only considers the input provided in the test, but also the results expected from the input by characterizing both the test tasks and the interaction of the test taker (Lewkowicz, 2000).

3. Importance of authenticity

Undoubtedly, authenticity is important to both language test designers and takers. This is a goal that test researchers and developers are pursuing together. According to Bachman (1990, cited by Brown, 2001), authenticity is important because it is a way to ensure that “language tests reflect language use in the target domain, and that their results are thus valid for application in that domain.” In the same way, Spolsky (1985, cited by Brown, 2001) declares that there is no adequate authenticity is a threat to generalization. In addition, Morrow (1991: 112) said authenticity bears “the overriding importance”.

Moreover, Bachman mentioned the importance of authenticity in his work with Palmer (1996): authenticity was crucial in that it could potentially influence test takers’ perceptions of the test and even their performance. Namely, it is favorable to make test takers perform excellently as well as to improve their positive response to the tasks. According to Bachman and Palmer (1996), it’s important for another reason it connects test tasks with the domain of generalization. Authenticity thus provides a measure of the extent to which the survey score is interpreted beyond performance in the domain of test language use or similarly in the domain of non-test language use. This links authenticity to construct validity.

4. Authenticity and construct validity

Construct validity is another key research area in language testing, which refers to meaningful and appropriate interpretations in terms of test scores (Bachman & Palmer, 1996). Generalization of survey score interpretation is essential to construct validity. Therefore, there is an inseparable relationship between construct validity and authenticity. Both authenticity and construct validity rely on how the linguistic competence of the test is defined. Authenticity is relative to the relevance of the test task in the TLU domain, and therefore to content validity. Also, it offers a basis for specifying the domain in which score interpretation can be generalized in addition to studying construct validity (Bachman & Palmer, 1996: 29).

5. Characteristics and Dimensions of authenticity

Since authenticity is pivotal to a test, it is necessary to learn about what features it bears. The study of authenticity is still the central part of the language test. In the past few years, language tests have been depicted as direct, performance, functional, communicative and authentic. The terms which are used to characterize authenticity in language tests frequently by test developers or researchers are “direct” and “performance” (Bachman, 1990). According to Bachman (1990), three basic elements can be applied to characterize authenticity: a view of language ability, or proficiency, as pragmatic ascription (Upshur 1979; see p 251 above, cited by Bachman, 1990); the reference to real-life performance as a criterion, and the belief that faces validity, content relevance, and predictive utility are sufficient bases to justify test use.
How authenticity can be measured or identified is also a question deserving to be considered. Elder (2016) notes that authenticity is a complex concept that is subjective in nature, and therefore it is difficult to define and measure. According to Brown (2001), it can be measured by at least four dimensions: test task, the interaction between tests and tasks (interactivity), the interaction between participants (paying attention to the interactional authenticity in language interviews as well as the co-constructed nature of ability) (McNamara, 1996; Young & He, 1998, cited by Brown, 2001), and test criteria as well as procedures (Actually, the relationship between the criteria used to judge performance and realistic judgments in a particular domain is the key to validity).

6. How to design a relatively authentic test

As an essential quality of a test, authenticity has to be considered in designing tests by test designers. Since authenticity contains two aspects: situational authenticity and interactional authenticity. How can a test both authentic in situation and interaction be designed? There are four interesting examples in Bachman’s work with Palmer “Language Testing in Practice” (Bachman & Palmer, 1996):

Example A is a test designed for new typists from a suppositional foreign country, some of whom are poor at English but good at typing given English tasks even documents written by hand. It is difficult for typists to speak or write in English independently. In this case, the test may be to require the examinees to type according to a prescribed handwritten document. If the examinees realize that they do not need to use a lot of English in daily work, they will perhaps be cognizant that the test is highly relevant to the job. Nevertheless, the test corresponds little to interactivity because the applicants are not asked to adopt the document in English. They just type in terms of the given letters. Thus, the test in this example is high in authenticity while low in interactivity (Bachman & Palmer, 1996).

The test in Example B is in the same situation as that of Example A. Suppose that the examinees have enough ability to talk in English about daily life, and they will be interviewed in English. If they are interested in the topics of the test, they may interact with the interviewer just like in a similar non-test interview. However, the purpose of the test is to choose new typists who just type according to the given English files. Thus, the task is rated to be “relatively low in authenticity” because of “the lack of relevance of the test task to the TLU tasks”; It is judged to be “relatively high in interactivity” especially when the interviewees are allowed to select topics according to their preferences and influence the structure of the interaction (Bachman & Palmer, 1996).

Example C is a hypothetic English vocabulary test for overseas students who are planning to enter an American university. The test format is to match words with their meanings in another column. The test result will be used to judge the students’ English academic reading competence. This test is supposed to be “relatively low in authenticity” because of its high restriction of language knowledge and its irrelativeness to the metacognitive strategies (Bachman & Palmer, 1996).

Example D is a test in which the testees are asked to play the role of salesmen to sell products. The test task includes a face-to-face conversation with a pretended customer. The testees have to decide what means they should use to attract the customer into the conversation. In this way, this test is rated as “relatively high in both authenticity and interactivity”. It is highly authentic because the characteristics of the TLU correspond with the characteristics of the test task. It is highly interactive because it involves “assessment”, “goal-setting” and “planning strategies”, and the level of language knowledge and topical knowledge involved is high (Bachman & Palmer, 1996).

The results of the four tests are shown in the following table (Bachman & Palmer, 1996):

<table>
<thead>
<tr>
<th>Interactivity</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authenticity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>Low</td>
<td>C</td>
<td>B</td>
</tr>
</tbody>
</table>

From the above examples and the table, we can conclude that an authentic test refers to a test with relatively high authenticity both in situation and interaction because situational authenticity of a test task is high while interactional authenticity is low or vice versa (Bachman, 1990). Spolsky (1985) once put forward a question: Why not construct a situation in which the testees cannot realize that they are being in a test and show their knowledge and skills authentically? The question reminds me of an interview with a language teacher in MQ when I was studying there. We were divided into four groups and were required to interview four teachers who did not know beforehand. After that, we had to write
a biography for the teacher we interviewed. The marks were only given for the biography, not for the interview, so all the students were interested in the interview because they were not pressured by grades and intended to know more about the teachers. Therefore, I believe that if only one language teacher was interviewed in the same situation, and the teacher judged the score according to each student’s language knowledge and communicative ability, it would be a relatively successful authentic test.

7. Problems in designing and implementing an authentic test

Despite the importance of authenticity, there are a lot of issues in designing and implementing authentic language tests (Spolsky, 1985):

Firstly, there are not adequate authentic materials available for language testing. Spolsky (1985) pointed out the problem in his work “The limits of authenticity in language testing”: The criteria of authenticity raise important pragmatic and ethical issues in language testing. The lack of authenticity in the materials used in the tests raises problems regarding the generalization of the results. He also proposed in the book that choosing an appropriate authentic material is always difficult. Functional materials are usually more authentic than those extracted from real situations. Authentic listening materials seem to be more difficult because they contain a lot of noises and irrelative materials which make it harder to be aware of what exactly is being examined (Ringeling, 1984, cited by Spolsky, 1985). Thus, test designers cannot use authentic materials in any examinations.

Moreover, a language test is inauthentic, artificial language behavior in itself because the purpose of a test is to investigate knowledge or competence in the language of a test taker while not just to know the answer to a question, it is impossible to design a completely authentic test. Above all, a testee needs to demonstrate his knowledge and ability, and therefore cannot be as completely truthful as non-test interactions. Just as Edelsky and her colleagues (1983, cited by Spolsky, 1985) argued: the process of a test is inauthentic no matter what you do. Therefore, no test is both authentic and natural in the use of language. It is only used to find out who is good at language and who is bad at language. In short, the test is inconsistent with authenticity, so it is impossible to develop an authentic test.

In addition, for those testees who are not ready to abide by the rules of a test or do not know those rules, the test results will be inaccurate and invalid to identify their knowledge and skills. According to Spolsky (1985), students in some places still maintain their traditional values and it is unacceptable to show their abilities in public. In such places, it is right not to answer questions or correct mistakes in classes. In this case, it is impossible to implement an authentic test.

Also, in his work, Spolsky introduces some ways to overcome problems. However, he recognizes that only part of them can be solved by authentic-seeming tasks (also called “planned encounter” by sociolinguists), which are performed based on the belief that the tester is interested in the information in the testees’ answers and the tasks are real questions. However, there is no escaping the fundamental problems. Therefore, he states that problems can only be solved if observers who observe patiently and sympathetically are willing to help (Spolsky, 1985).

8. Relativity of authenticity

Although many researchers have studied authenticity for years, no one can design an authentic language test. Breen (1985, cited by Lewkowicz, 2000) points out that the goal of language learning is to be in a position to comprehend the meaning of the text, so texts aimed at achieving this goal can play a role in teaching. He claims that authenticity is a rather complex concept and that it is difficult to separate which is authentic and which is not because of its relative nature. Similarly, Bachman (1990) also considers the complexity of authenticity and points out that both situational and interactional authenticity are relative rather than absolute. The situational authenticity of a test task may be high while interactional authenticity is low, and vice versa. Thus, the correct statement is “relatively more” or “relatively less” authentic or interactive, rather than “authentic” and “inauthentic”, or “interactive” and “non-interactive”.

9. Conclusion

In summary, as a crucial attribute of a test, authenticity has been debated both in the field of general education and applied linguistics. Despite the positive attitude of test designers and researchers, they have to face some limitations when designing and implementing an authentic test. Such as the lack of adequate authentic materials, the artificial nature of the test, the negative reaction of testees, the relativity of authentic tests, etc.

In conclusion, while designing and implementing an authentic test is the common goal for both test designers and developers, it is impossible to design a truly authentic test because of so many negative issues. Test researchers must strive to design a relatively authentic test.
References