

Authenticity in Language Testing

Yunai “Amy” Wang

Abstract

There has been an increase in awareness and research on authentic tests used to measure language learners’ ability to use the target language in real-life situations. In order to serve this purpose, test designers should place value on authenticity in tests. This paper begins with a brief overview of authenticity in language testing. This is followed by a discussion of several characteristics of authenticity in particular types of tests. Finally, some limitations and problems in the implementing of authenticity in language testing are presented.

Introduction

Being a passive English learner for many years and having taken a great number of tests, I know the effect that the language testing has on both language teaching and learning. In studying principles of good language assessment, I learned to distinguish between different forms of assessment. I realized that besides traditional forms of testing, there are alternatives that teachers could use to assess students’ language ability, such as performance tests, self-assessment, and portfolios. Among various forms of alternative assessment, authentic testing has become the trend in recent years. In completing authentic tasks, language learners use what they have learned in class in real-life situations, which is the main objective of language learning. There are several questions regarding the authenticity in testing that I would like to explore in this literature review. I would like to know about authentic tests and authenticity in language testing in general as well as the characteristics of authenticity and its limits. I hope that in my future teaching, I can make tests that provide students with more opportunities to perform successfully in real life situations, which is what they need.

What is Authenticity in Language Testing?

In order to make a test that could effectively measure learners’ language ability and proficiency, the experts frequently mention these factors as important: validity, reliability, practicality, and appropriation. In addition, authenticity is another important factor that test makers need to focus on and

consider in the process of developing the tasks. The centrality of authenticity in assessment can be seen in the following statement by Bo (2007), “authenticity has become established as a central concern in test design and test validation” (p. 31). Since the purpose of language testing is to “reflect the use of language in the ‘real world’”(Spence-Brown, 2001, p. 463), authentic tasks have become one of the most important components of tests that aim to measure language learners’ ability to apply what they have learned from class to real world situations. Leung and Lewkowicz also suggested that authentic test tasks should be made based on the real-life resources which are close to the test takers’ expectations when they actually do the tasks outside the classroom or test situation (2006). If the test takers are capable of implementing the tasks successfully, such as having a job interview or writing an advertisement for the local newspaper, we could say that the tests are authentic because they bring in the learners’ needs and the events they might encounter in real life. Thus, authentic test is considered “true assessment of performance” (Wiggins, 1998, p. 21), as opposed to traditional tests, such as entrance examinations with multiple choice, gap filling, and true or false questions and rigidly formatted essay writing. In this sense, an authentic test is designed for testing in language for specific purposes, which can be distinguished from general purpose language testing (Wu & Standsfield, 2001). Simply put, an authentic test is “one that reproduces a real-life situation in order to examine the student’s ability to cope with it” (Doye, 1991, p. 3).

In addition to the link between the test situation and the real world situation, authenticity in language testing also involves interaction and communication. Bo (2007) stated that in authentic tasks, meaning and communication is more important with respect to the validity of the test, especially when it comes to productive skills, which require test takers' interactive and communicative ability. In this view, authenticity is more focused on the connection between the test takers and the test itself, and it depends on how the individual interprets the test tasks, which "will affect both the validity and reliability of test scores" (Bachman, 2007, p. 10). Therefore, in developing such tests, apart from the measurement of the candidate's linguistic ability, test makers need to be familiar with the fundamental meaning of test materials, and their tests need to reflect the communicative competence of the learners as well (Bachman, 2007).

Achieving Authenticity in Testing

After understanding the definition of authentic test, it is important to know what the test might look like, what characteristics it has, and some of the factors involved in developing it. The examples below provide some insights into the implementation of authenticity in language testing.

Content Authenticity in the Test **Authenticity in direct testing**

In order to guarantee content validity, in the development of the test, test makers should consider the purpose of the test and the particular target group of the test in order to make sure the test serves as an assessment tool for language for specific purpose (Wu & Stansfield, 2001). This means that test makers must understand target language use in real life conversation, for example, "production responses" in the work environment (Chun, 2006, p. 295). Wu and Standfield (2001) also stated that job analysis of roles and settings should be the basis of test content. By doing tasks related to jobs, test-takers are asked to use the target language and rely on their real world knowledge to

complete the tasks that are "appropriate to various contexts" (McNamara, 1996, p. 450).

For example, in the Listening Summary Translation Exam in Taiwanese (LSTE/T), the tasks given to the test takers take two forms, a face to face interview and a telephone interview. The test was designed for the US government to select qualified linguists for work in law enforcement agencies. In order to maintain content authenticity, in the development of the tasks, the test designers tried to make the tasks as close as possible to the situation in the real world, which was to understand the conversation between criminals. In the test, there are two sections: fifty multiple choice items based on the information in recorded conversations and a summary writing task based on the listening materials. Here are some sample multiple choice questions:

1. Who are these people? (old/young; male/female; well educated/not well educated)
2. How do the participants talk? (vulgar, rude, polite, etc.)
3. What kinds of emotions do they display? (anger, frustration, concern, self-confidence, happiness, business-like, secretive) (Wu & Stansfield, 2001, p. 205)

Summary writing items:

4. How often do they use words or phrases from other dialects or languages, like Mandarin, Cantonese, English or other languages?
5. Can you give me some examples of the way they demonstrated such emotions?
6. In the conversations you listened to, do the speakers ever talk about problems they are having? What kinds of things do they say? (Wu & Stansfield, 2001, pp. 205-206)

In this particular example of testing language for specific purposes, the test-takers' performance was supposed to reflect their actual competence, and that is what the test designers expected. Even though the researchers claimed that the multiple choice

items could “assess the understanding of the specific information” (Wu & Standsfield, 2001, p. 197), I do not think they met the requirements of authenticity, and the researcher acknowledged that it was “more difficult to achieve this goal” than s/he had realized it would be (Wu & Standsfield, 2001, p. 197). In my view, the items should be presented orally, which could measure the test takers’ competence more directly.

Authenticity in indirect testing

When it comes to authentic tests, people often think of speaking tests, which require test takers’ actual performance. However, authenticity can also be reflected in the indirect tests such as reading comprehension and the accompanying writing tasks as well.

Traditionally, second language reading comprehension tests do not measure students’ actual ability in understanding the texts due to the use of multiple choice items, which creates a gap between test situations and real-life situations. In order to make tasks more authentic and able to measure test-takers’ competence in solving problems in life, reading should be considered as an “interactive process between the reader and the text” (Cardoso, 1998, p. 52). To this end, traditional multiple choice items were replaced by more complex essay questions in the English test portion of a university entrance exam in Brazil. Since the aim of the test is to select applicants who are able to express themselves and demonstrate their knowledge to prove their capabilities effectively in college, the test designers viewed language proficiency as a more important factor than language knowledge (Cardoso, 1998). Through writing the answers, the applicants could show their capabilities of comprehension, critical thinking, and expression. It also gave them more opportunities to “show their active relationship with knowledge through their writing” (Cardoso, 1998, p. 53). In addition, using authentic materials in the test, such as newspaper articles and advertisements, also contributed to its authenticity. For example, one question was, “Explain the title of the advertisement.” The question not only assesses the students’ reading ability to make connec-

tions within the text itself, but it also involves the relationship between the individual’s background knowledge and the text.

Similarly, Bo (2007) also mentioned the value of authentic language in tests. As an example of authentic tasks, he mentions the College English Test-Band Four and Test for English Major-Band Eight tests in China, which are public examinations in English to measure students’ language ability based on the syllabus. There are easy-to-understand 10 short daily conversations in the listening comprehension part and the interesting and realistic topics in reading comprehension part in both tests. He suggested that the testing materials should be “quite familiar to the test-takers” (Bo, 2007, p. 31). It should be relevant and understandable to the test takers so that they can combine their prior knowledge and the given materials to process and answer the questions. In addition, in his analysis of the two tests, Bo (2007) also stated that “productive skills involve test takers into [sic] interaction and communication” (p. 32). In other words, in the development of tests, test makers have the responsibility to make sure those tasks “are as close as possible to the tasks in real life which test-takers are expected to perform” (Wu & Standsfield, 2001, p. 203).

The use of different languages in authentic tasks

Besides using the target language in tests, surprisingly, test-takers are sometimes allowed to use their native language to complete tasks. The use of the test takers’ native language can sometimes be found in the reading comprehension items and speaking tests. In a university exit exam in Brazil, the instructions for the reading comprehension section are in the students’ native language and so are the answers given by the students. The test contains authentic texts such as newspaper articles and poetry, but the instructions and questions are in Portuguese. The use of students’ native language was viewed as one of the merits of the tests. “The students are not asked to write in the L2.....that none of the other language skills are required for students to complete the tasks except for their reading competence”

(Cardoso, 1998, p. 55) Even if the target language in the answers is expected in most of the test situations, the use of the students' native language does not affect the authenticity of the test. In the real world, it is realistic to cope with the problems in L1 when the problems are presented in L2. Hence, the students are able to use the native language to express their ideas and to show their comprehension of the reading tasks. Likewise, in the Listening Summary Translation Exam in Taiwanese, "one speaker may ask a question in Taiwanese, but the answer may be given in Cantonese, Mandarin, English or even Vietnamese" (Wu & Standsfield, 2001, p. 190).

Interaction and communicative authenticity in oral tasks

In order to make sure that students are capable of using the target language in real situations, the tasks should be given in circumstances which involve interaction and communication between the interlocutors. Interaction "influences the test process" and also combines the general cognitive capacity such as "grammar, vocabulary and sociolinguistic appropriateness" (McNamara, 1996, p. 450).

In one interview task of a Japanese test for students at an Australian university, the students of Japanese were asked to interview native speakers outside of the classroom in order to write a report to show their language level of proficiency. The task contained authentic interaction between participants. In order to collect the information for the written report, students were required to tape their communication with native Japanese speakers, which provided them with opportunities to practice the target language in a real life situation. In this way, students were highly motivated because they got to choose their own topics, which "increase[ed] authenticity and motivation" (Spence-Brown, 2001, p. 467). Interaction and communication is one of the components of authentic tasks, according to Bo (2007). It improves the learners' productive skills, competence and confidence. It also creates a "great harmony between testing practices and the way language was being

taught" (Leung & Lewkowicz, 2006, pp. 215-216). Therefore, by conducting authentic tasks, test-takers have the opportunity to utilize their knowledge in practical settings. Gradually, through preparation and practice, this communicative skill is likely to be internalized, which is beneficial to the students' improvement of language competence.

Limitations of Authentic Testing

So far, I have provided an overview of how authenticity works in several test tasks. However, there are some limits and problems that we as language teachers need to be aware of. These include the following:

1. It is impossible to have a complete congruence between the test situation and real-life situations (Doye, 1991).
2. Test takers' characteristics and attitudes toward the tasks may affect performance and the score (Lueng & Lewkowicz, 2006).

In an attempt to make the test tasks mirror exactly the situation in real life, test designers have to consider the test takers' specific needs. If they do so, the degree of authenticity that can be achieved realistically is affected because it depends on the particular purpose for which the test is designed (Bo, 2007 & Spence-Brown, 2001). However, the basic problem for language testing is that tests are created in artificial contexts (Wilfried & Jozef, 1997), especially tests related to job-relevant interviews. Because of the "restrictions imposed on language testing, no matter how much like real-life tasks seem to be, they are not real communicative tasks" (Bo, 2007, p. 32). Therefore, Bo acknowledged that "authenticity is a relative term" (p. 28).

Similarly, when doing authentic tasks, test-takers' performance and behaviors do not always reflect exactly what they do in the real world. Different test takers seem to interact with the test in different ways, and sometimes the test does not elicit the language sample that test makers expected (Bachman, 2007). One example of this can be seen in the case of the interview tasks that Australian university students have for Japanese written report cited above. In the

process of preparing for the tape recording, each student adopted different methods, trying to show as much language proficiency as possible. However, due to the students' differing attitudes and interpretations of the task, they performed differently in the discourse task. Two out of four students emphasized authenticity when they reflected on their understanding of the test's objectives. As a result, they tried to engage themselves with the task as authentic interaction, which resembled "real" conversation (Spence-Brown, 2001). For the other two students, they placed more priority on assessment with good grade than the chance to practice the target language. And among them, only one student's framing of the task "matched most closely that of the task designers" (Spence-Brown, 2001, p. 472). Although there are only four students mentioned in the case study, I believe they are probably quite representative of students' perception in general. In order to increase reliability, I think clear explanation of the test's objectives should be provided to the candidates before the test.

References

- Bachman, L. (2007). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bo, J. (2007). An analysis of authenticity in CET-4 and TEM-8. *Sino-US English Teaching*, 4 (2), 28-33.
- Cardoso, R. (1998). Authentic foreign language testing in a Brazilian entrance exam. *Texas Papers in Foreign Language Education*, 3 (2), 51-70.
- Chun, C. (2006). Commentary: An analysis of a language test for employment: the authenticity of the PhonePass test. *Language Assessment Quarterly: An International Journal*, 3 (3), 295-306.
- Doye, P. (1991). Authenticity in foreign language testing. Retrieved October 16, 2007, from http://www.eric.ed.gov/ERIC-Docs/data/ericdocs2sql/content_storage_01/0000019b/80/12/ed/39.pdf.
- Leung, C. & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: language testing and assessment. *TESOL Quarterly*, 40(1), 211-234.
- McNamara, T. (1996). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18 (4), 446-466.
- Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing*, 18 (4), 463-481.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. SF: Jossey-Bass.
- Wilfried, D., Jozef, C. (1997). Competence and performance in terms of content validity in productive language testing. In M. Gardenghi & M. O'Connell (Eds.), *Prijfen, testen, bewerten im modernen Fremdsprache-nununterricht* (pp. 25-36). Frankfurt: Verlag Peter Lang.
- Wu, M., Stansfield, W. (2001). Towards authenticity of task in test development. *Language Testing*, 18 (2), 187-206.

Conclusion

Authenticity is one of the crucial components that test designers should include if the goal of testing is to reflect test takers' actual language competence and abilities. There are two main aspects to be considered regarding authentic tests: the connection between the test situation and real life situations and the interaction between the test taker and the tasks. Furthermore, the test also needs to guarantee content validity and provide test takers with as many opportunities as possible to carry on the tasks successfully and effectively. Alongside the advantages, authentic tasks also have some limitations and problems that might affect test takers' performance. Test designers should understand test takers' goals of learning a language, clarify the objectives of their tests, investigate the real situations on which their tests are based, and experiment with the authenticity and practicality of the tests in order to avoid the negative backwash caused by assessment tasks.

[This page is intentionally blank]