

ALTERNATIVE ASSESSMENTS IN READING : MAKING THE TEST MORE AUTHENTIC

Wiyaka
English Department- IKIP PGRI Semarang
Email : alex_wiyaka@yahoo.com

Abstract

This paper tries to evaluate the existence of reading test format used in national exam which has been applied for many many years in our schools. As we know that under the reasons of practicality and limited budget from the government, a multiple choice type is used to assess the students' language competence. Logically, language competence is not only assessed through recognition test, but also production test. In that way then it is regarded important to provide test types other than traditional test in the hope that the test does not only function as a way to measure the product of learning, but also the process of learning. Some alternative assessments will be discussed and offered to the language practitioners.

Key words : *assessments, authentic*

INTRODUCTION

Teaching English at schools always deals with at least three components; materials, methods, and assessments. Materials correspond with what to teach; this usually becomes the main concern of the teachers and it is very logical since the first thing the teacher has to consider in mind is what suitable materials should be prepared on the bases of standard competences stated in the curriculum. Methods deal with how to teach. It is a matter of the techniques or approaches on how the learning outcomes are achieved by the students. The third component, assessment, deals with how the teacher monitors the learning progress and achievement of the students during the teaching learning process or at the end of the term.

Assessment or evaluation or test is basically a part of teaching. It cannot be separated from teaching in the sense that there is no teacher with speciality in teaching but not in testing or the vice versa. In his book Brown illustrates how tests and assessment are integrated in teaching (see Brown 2004 p.5). Test is subset of assessment. Brown says that assessment is ongoing process. It is done during the process of teaching and learning. It can be formal or informal. Any teacher's remarks or instructions to check the students understanding or perception is classified into assessment. While test is said to be very technical in terms that it is an administrative procedure that must be prepared by the teacher prior to the schedule of the test.

It is clear that within teaching activities there assessment which is defined as an ongoing process done by the teacher in observing and measuring the students' performance during the whole process of learning. This includes not only formal tests, but activities such as students' response to questions, students comments, students' inquires etc. In the technical terminology, assessment differs from tests in that the latter is prepared, scheduled and administered sometime in advance so that the students know their performance are being measured and evaluated.

Tests must be based on the learning outcomes which are stated in the curriculum. A test must measure the standard competences and basic competences. To see whether the standard competences are achieved or not one must use a kind or kinds of instrument either it is a test or none-test technique. In the school practices one instrument familiar to teachers is test. However, not many teachers are aware of the test constructions. They usually make use of the tests form the provided samples found in the texts books or other sources, taken for granted. Or, it is because of the fact that at the end of the year the school teachers are not required to write or design the test by themselves as the test is made by a team assigned by the local government.

The problem become more serious when it is related to national examination in which the test items are prepared by a team from the ministry of education. It shouldn't have been a problem if the teacher could identify the specifications of the test items related to the language competences required by the curriculum. What is usually done by teachers was giving students exercises from day to day or for months prior to the examination. Learning is not a process of acquiring the competence but making preparation for the test. This becomes worse when the test items are all in multiple choice in which the students are trained to select the correct answer. This means that it tests the knowledge "about language", not "the use of language".

THE EXISTING NATIONAL EXAM : SOME CRITICMS

When we look back to the format as well as the contents of national exam especially in English subject for SMA level, there are some shortcomings to be addressed. First, among the four language skills or language competences that should be taught and then assessed, only listening and reading are tested. Listening test consist of 15 items (30 % of the total number) and reading test comprises of 35 items (70 %) on reading comprehension. All take the form of multiple choice format. So, it is a kind of recognition test rather than production test. The question might arise whether multiple coice form is the only test mode to measure

the students' listening and reading competencies. Does recognition type of test really measure the real tasks of language use in real life? In other words, the validity of the test construction is questioned.

The fact that the two other language skills, i.e. speaking and writing are not tested also reflects the incompleteness of the test coverage. These two skills are not tested due to some reason. First, the budget needed for administering those two is quite high since testing the two skills cannot be done through multiple choice format which means impractical in one side. It must be done through performance test for speaking and writing. We can estimate how many test raters or test supervisors are needed for such performance test. It must require a lot of fund from the government to support the test supervision as well as the rating or the scoring of the result. The second problem is on the readiness of the test raters. Do they have enough expertise or skill to evaluate speaking performance? We frequently hear the bad news of the English teachers' competence. The result of Teacher Competence Test (UKG) recently held by the ministry of education reveals the the average achievement score gained by the test takers is only 40 out of 100. It is very disappointing. Then how can such teachers become good raters for their students speaking competence. If they are to be trained several weeks to become competent raters, then how much expense the government should provide to support the training of a hundred thousands of the teachers. So the absence of testing speaking and writing is due to the financial problem and the human resource availability. This is of course just controversial to the required competences listed in the curriculum in which the competencies of speaking and writing should be fostered in the students learning and consequently on the evaluation.

The literature also presents an array of negative criticism with regard to the 'washback effects' or consequences of high-stakes standardised tests like UN (Ujian Nasional – National Exam) on a number of levels. According to Tsagari from Center for Research in Education Lancaster University, high-stakes exams present some bad consequences as the followings.

1. Curricular level

Critics of high-stakes tests attest that these are responsible for narrowing the school *curriculum* by directing teachers to focus only on those subjects and skills that are included in the examinations. As a consequence, such tests are said to “*dominate and distort the whole curriculum*”

2. Educational level

Critics also point out that high-stakes examinations affect

- a. *the methodology teachers use in the classroom*, i.e. teachers restrict the methods they use and employ various exam preparation practices (also known as “coaching” or “cramming”) at the expense of other learning activities which do not always contribute directly to passing the exam
- b. *the range, scope and types of instructional materials teachers use*, i.e. highstakes exams gradually turn instructional materials into replicas of the actual examination papers
- c. *students’ learning and studying practices*, i.e. in high-stake examination contexts students tend to adopt ‘surface’ approaches to learning as opposed to ‘deep’ approaches. As a result, students’ ‘reasoning power’ is impeded, rote memorisation is encouraged by concentrating on recall of isolated details and students resist attempts to engage in risky cognitive activities which can prove both effective and potentially beneficial for their future improvement

3. Psychological level

Furthermore, high-stakes standardised tests are also said to have undesirable effects on:

- a. *students’ psychology*, i.e. it is believed that the role of the students in contexts where high-stakes tests are introduced is that of passive recipients of knowledge and their needs and intentions are generally ignored. High-stake tests are also said to have detrimental consequences on students’ intrinsic motivation, self-confidence, effort, interest and involvement in the language learning experience and induce negative feelings in students such as anxiety, boredom, worry and fear, which, according to the literature, are not conducive to learning.
- b. *teachers’ psychology*, i.e. it is argued that the dictates of high-stakes tests reduce the professional knowledge and status of teachers and exercise a great deal of pressure on them to improve test scores which eventually make teachers experience negative feelings of shame, embarrassment, guilt, anxiety and anger.

In addition to the above, it is also argued that teacher-made tests, if used as the sole indicators of ability and/or growth of students in the classroom, may generate faulty results which cannot monitor student progress in the school curriculum. It is also believed that the use of tests in classroom settings tends to overemphasise the grading function more than the learning function of the language learning process. As Black and William (1998) point out, in such contexts there is a tendency to use a normative rather than a criterion approach to assessment which is likely to create competition between pupils rather than personal

improvement leading to de-motivation and making students lose confidence in their own capacity to learn. In addition, it is also said that teachers do not generally review the assessment questions or tasks they use in their classroom tests and do not discuss them critically with peers. As a consequence there is little reflection on what is being assessed (Black and Wiliam, 1998). Teachers, according to Black and Wiliam, also do not trust or use their test results as these do not tell them what they need to know about their students' learning and appear to be unaware of the assessment work of their colleagues, too.

1. Alternative Assessment : A way to make tests more authentic

There is no single definition of '*alternative assessment*' in the relevant literature. For some educators, alternative assessment is a term adopted to contrast with standardised assessment, e.g. professionally-prepared objective tests consisting mostly of multiple choice items. Experts look at alternative assessment in more general terms. For instance, Hamayan (1995) sees that alternative assessment "*refers to procedures and techniques which can be used within the context of instruction and can be easily incorporated into the daily activities of the school or classroom*" (ibid:213). To this Smith (1999) adds that "*alternative assessment might take place outside the classroom or even the institution at various points in time, and the subjects being tested may be asked to present their knowledge in various ways.*"

While Alderson and Banerjee (2001) as quoted by Tsagari provide the following definition:

'Alternative assessment' is usually taken to mean assessment procedures which are less formal than traditional testing, which are gathered over a period of time rather than being taken at one point in time, which are usually formative rather than summative in function, are often low-stakes in terms of consequences, and are claimed to have beneficial washback effects.

It is clear from the above statement that alternative assessment is appropriate to assess the process rather than the single product of learning. It is not mainly used to justify whether a student will pass or fail the exam.

The term *alternative assessment*, and particular testing practices associated with it, have recently come into vogue in language testing. The movement is directed at establishing qualitative, more democratic, and task-based methods of evaluation in testing a learner's language proficiency (Brown and Hudson 1998). It contrasts with traditional methods of testing by involving the learners in the evaluation process, and having the

tendency to locate evaluation in a real-life context and, as result of these two features, being longitudinal. Thus, the insights emanating from these methods, alongside being used for decision-making about the future of learners, contribute to and furnish additional instructional purposes. As McNamara (2000 in Tsagari) points out: “*This approach stresses the need for assessment to be integrated with the goals of the curriculum and to have a constructive relationship with teaching and learning*”. The procedures used within this paradigm include checklists, journals, logs, videotapes and audiotapes, self-evaluation, teacher observations, portfolios, conferences, diaries, self-assessments and peer-assessments (Brown and Hudson 1998). These procedures have been diversely called alternative or performance assessment as opposed to traditional assessment techniques such as multiple choice, cloze test, dictation, etc.

While the new movement promises more humanistic and rewarding methods of testing and thus has a lot to offer, most teachers are not quite familiar with the new concepts and practices within the emerging paradigm. To enlighten the views of interested teachers, it can be a good start to answer a basic question about the so-called alternative methods of testing which may have occupied their minds. This question is concerned with the relationship of these *other* methods with the traditional methods normally used within classrooms.

2. Alternative Assessment in Reading

A test is made to measure certain objectives. Most language tests measure one's ability to perform language, that is, to speak, write, read, or listen to a subset of language. On the other hand, it is not uncommon to find tests designed to tap into a test-taker's knowledge on a language such as: defining a vocabulary item, reciting a grammatical rule, or identifying a rhetorical feature in written discourse. Performance-based tests sample the test-taker's actual use of language, but from those samples the test administrator infers general competence. A test of reading comprehension, for example, may consist of several short reading passages each followed by a limited number of comprehension questions—a small sample of a second language learner's total reading behavior. But from the results of that test, the examiner may infer a certain level of general reading ability.

The tasks of the reading tests so far are focused on answering the questions of the text. And most commonly it is done through multiple choice format. It seems monotonous to students despite the fact that there are some disadvantages of multiple choice of test. For example students can “guess” the answer, or students can cheat the

answer. Moreover, such tests do not reflect the real-world life communication as suggested in the curriculum. Curriculum 2006 highlights the communicative competence as the central objective of English language teaching in our schools. So, it is an irony to see that there is a contradiction between the required competence and the tasks of the test. This is in contrast with what Fulcher (2007 p.63) says:

...only tasks that mirror language use in the real world should be used in communicative language tests, reflecting the actual purposes of real world communication, in a clearly defined contexts, using input and prompts that not had been adapted for use with second language speakers.

There are some alternative assessment other than multiple choice that can be done by teachers, especially when the goal is to enhance students communicative competence. By using alternative assessments the students are forced to maximize their language exposure in using language actively. Some types of assessment that can be applied in reading are the followings:

1. Gap-filling formats (rational cloze formats)

It is similar to a cloze test but differs in that gap-filling measures (rational cloze formats) targets specific words purposefully (e.g., prepositions, verbs) rather than delete every seventh word (for example). However, even with gap-filling formats, a reading measure should not ask students to fill in words (as a production task) that they do not know or have not already seen from reading a text beforehand (unlike short-answer formats in which students have read a non-mutilated text beforehand).

2. Text gap format

Text-gap formats involve the moving around of whole sentences or paragraphs, or the selection of the right space in the text to supply a sentence or paragraph. Text-gap formats can be tricky when multiple gaps are created and a list of sentences or paragraphs is provided to insert in the correct spaces. These formats amount to a type of multiple matching task. Choosing from a heading bank to label identified paragraphs is a similar type of task. The strength of these types of tasks is that they call on knowledge of discourse signals and discourse structuring to be answered successfully. They require several comprehension skills for appropriate task completion.

3. Free recall format

Free-recall formats simply ask test takers to make a list of ideas they remember from a text they have just read. These responses are matched up against a list established

by the test maker. Summary formats can be straightforward though difficult to score. Alternative summary formats can include, for example, choosing the best from among three summary options and identifying the weaknesses of unacceptable options.

4. Project-performance evaluation

Project-performance evaluation is a newer task format that evaluates test takers as they read texts and then perform in groups to carry out a larger project. It is an interesting option, but is problematic on several validity grounds (giving individual scores based on group interactions and a holistic task)

5. Open-ended questions

It is much more challenging than multiple choice format. Test takers are free to compose own sentences related to the ideas from the given text. It is productive rather than recognition test. To make it more effective teachers can construct questions requiring high order thinking.

6. Writing Sample

This type is much more productive skill than the other formats. The students are fostered to reconstruct the text they are reading in a personal written report. They can use their creativity in language to paraphrase the text. This can be more meaningful when different students get different texts so there will be no cheating from the peers.

7. Text retelling

This is similar to writing sample except that the response is in the spoken form. Because it is spoken, the students get opportunity to tap their own stock of vocabulary as well as grammatical knowledge to communicate ideas related to the text they have read. In this way, reading competence is measured through speaking performance, so in one activity the teacher can generate two competences.

CONCLUSION

The alternative assessment paradigm, as discussed in this paper, is seen to give alternative methods to assess students reading competence. Reading is surely a receptive skill, but it does not mean that it must be evaluated by using receptive test format like multiple choice. English teachers should develop instruments to enhance the quality of their teaching, so they are not dictated to focus only on the preparation for the final exam which seems to be less beneficial

to the teaching itself. Assessment is an essential part of the learning process. However, further theoretical and empirical work needs to be done to examine alternative assessment practices in depth. For example, we need to reconceptualise alternative assessment and its relationship to standardised testing, to understand how the aspects of alternative assessment are actually accomplished in classroom interaction. Teachers should look back the standard competences mandated in the curriculum to see whether those competences, especially reading competences, could be attained by using alternative assessments. Teachers should also find sources of information on the effectivity of using alternative assessments on teaching and learning. There are some research studies on this that can inspire teachers before any definite conclusions about its positive effects on teaching and learning. If possible teachers can conduct classroom research on this interesting topic.

REFERENCES

- Bachman, L. F. Palmer, A. 1996. *Language Testing in Practice*. Oxford University Press
- Brown, H. Douglas. 2004. *Language Assessment, Principles and Classroom Practices*. San Francisco: Longman
- Fulcher, Glenn and Davidson, Fred. 2007. *Language Testing and Assessment*. New York: Routledge
- Hamayan, Else V. 1995. Approaches to Alternative Assessments. *Annual Review of Applied Linguistics*. Cambridge University Press
- O'Malley, J. M. and Pierces, L. V. 1996. *Authentic Assessment for English Language Learners Practical Approaches for Teachers*. New York: Longman
- Pusat Kurikulum. 2006. *Kurikulum Bahasa Inggris 2006*. Jakarta: Depdiknas.
- Tatsuki, Donna. 2006. Authentic Communication: Proceeding of the 5th Annual JALT Pan-SIG Conference. Shizuoka, Japan: Tokai University College of Marine science (p. 1-15). <file:///H:/authenticity.htm> accessed on June 21, 2012
- Tsagari, Dina. 2004. Is there life beyond language testing? An introduction to alternative language assessment. CRILE working paper no 58. Lancaster University www.carla.umm.edu/assessment/MLPA/CoRA.html). Accessed on November 18, 2012
- www.cambridge.org. Reading in a 2nd language.17_pdf. Accessed on 3 Dec 2012.