

タイトル	Communicative Language Testing
著者	ZEFF, B. Bricklin
引用	北海学園大学学園論集, 133: 29-36
発行日	2007-09-00

Communicative Language Testing

B. Bricklin ZEFF

Abstract

Some students have a difficult time learning a language to a level that allows them to apply that knowledge in a simulated real-life situation. Communication Language Testing (CLT) assesses a student's performance in a simulated situation such as a mock interview (where the student is asked a number of questions that exemplify good comprehension) or a testing task that would grade good or *competent* social and speaking ability. These tasks involve certain speech acts that are deemed valid and exemplify a standard with respect to certain specified abilities such as asking directions, describing a friend, or ordering food in a restaurant, in a "real" setting. In this paper, I will discuss the various tests used in language classrooms in Japan and how they compare to current practices in CLT. I will also assert that Communicative Language Ability (CLA) is a valuable language learning methodology and should be incorporated in the foreign language curriculum throughout Japan.

Introduction

"What does it mean to know how to use a language?" asks Spolsky (1985 p. 180). After living and teaching English in Japan since the late 1980's, I am often confronted with this question. Even though my students have been studying English for seven years, most are incapable of answering a simple question such as, "What do you do?" This is simple question to a native speaker. In this context, a possible answer might be, "I am a student." Another response could involve the construction of simple tense sentences to describe work or school experiences, such as "I go to school" or "I work part time at a convenience store." But to my Japanese students this question proves troublesome. Though they know how to construct a proper sentence, in this situation my students can barely distinguish what tense to use in their responses. In a classroom setting, however, they passed a variety of grammar tests where they were able to pick out, from three or four possible answers, the correct sentence.

But when asked this simple question in this context, they freeze.

In discussing their inability to transfer their knowledge of sentences from a test situation into a conversational setting, most feel that the opening “What do you do?” is confusing. The tense form is ambivalent. They are accustomed to the text-based use of verbs and question formation. You would rarely experience this direct question in text form, except in a conversation test. A native English speaker, on the other hand, would encounter this type of question in multiple contexts: home, school, and work. Does the failure to make sense of a simple question in English conversation show a lack of understanding of how a language is used?

— Testing Approaches —

According to Spolsky, “language testing and linguistic theory must try to define language knowledge and use” (Spolsky, 1985 p. 180). Spolsky describes three main approaches: structural, functional, and general proficiency.

The Structural Approach assumes that knowledge can take the form of a grammar or structural description of a language and forms the basis of discrete point tests. The Functional Approach assumes that the nature of language knowledge is best captured by listing the various uses for which it can be put and is embodied in the communicative competence model, the notional-functional curriculum, and the interest in teaching and testing pragmatics. The General Proficiency Approach is based on the concept that individuals vary in possessing measurable amounts of an indivisible body of knowledge. This approach underlines arguments for a general factor underlying batteries of tests or for the trait measured by certain test methods like a cloze, the fill in the blank exercise to test vocabulary and comprehension of words and expressions in the context of a dialog or essay. (Spolsky 1985 p. 180)

“The theoretical strengths and weaknesses of these approaches and the impossibility that any one is completely correct, forces us to consider all three as basic to testing.” Spolsky concludes that, “anybody who knows a second language must be assumed to have all three kinds of knowledge, so that we can only achieve the full picture of language proficiency if we use many different measuring methods and know what trait is being tapped by each test” (Spolsky, 1985 p. 180) This is to say that the testing process is not a one shot know all system. There should be ongoing testing and evaluation of the second language learner utilizing a variety of methods for determining fluency.

Discrete point tests break the elements of language into parts and tries to test them separately with little consideration to the ways those elements combine to facilitate a larger context of communication. (Weir 1990 p. 2) Most Japanese students have learned language within the Structural Approach. They have spent the seven required years of language study taking discrete point tests almost exclusively. As a language teacher fully entrenched in the Japanese educational system and faced with the daunting task of helping these students realize their English ability, I find that giving them the other parts of the puzzle, are paramount to my course design. The folly of limited testing is that in language *the whole is greater than the sum of its parts*.

Another form of testing that is used widely in teaching language skills in classrooms around the world is the integrated test. According to Oller (1979)

The concept of the integrative test was born in contrast with the definition of discrete point tests. If discrete items take language skills apart, integrative tests put it back together. Whereas discrete items attempt to test knowledge of language one bit at a time, integrative tests attempt to assess a learner's capacity to use many bits all at the same time, and possibly while exercising several presumed components of a grammatical system, and perhaps more than one of the traditionally recognized skills or aspects of skills. (Oller, 1979 p. 37)

From the late 1990s, the Education Ministry in Japan has discovered its shortcomings with regard to language education. This has led to a big shift in curriculum development and methods of testing used in public and private schools. But there remains a large number of students in colleges and universities today who lack even basic functional skills. These functional skills are embodied in communicative competence models, according to Spolsky (1985), and I believe these skills should be reflected in communicative language testing in Japan.

Weir (1990) describes a difference between what it is to test a student's performance in a language and what it is to test language competency. Performance testing involves testing a student's performance in an isolated situation (Weir, 1990 p. 12). When we consider the ability to deal with language in a variety of other situations," competence as well as performance would seem to be involved" (Weir, 1990 p. 9). Performance language testing asks a student to perform certain tasks such as job theme conversations. This might involve a student taking on the role of a hotel clerk and having an unscripted conversation with a test giver playing the role of a customer. The goal of this exchange would be for the student to

navigate through the conversation giving information, responding to questions, and possibly dealing with the unpredictability of certain forms of communication that a hotelier might experience. According to Alderson “language in use is unpredictable and so should our tests be” (1981, p. 3). How this student would be able to handle a conversation with an acquaintance about the political situation in the Middle East, or politics in general could not be assessed because it would involve forming opinions and making inferences that a conversation serving a customer would not necessarily address.

Tests including discrete point tests and integrative proficiency tests tend to test a student’s knowledge of language and structures in a static environment, with static tools, (i.e., a written test, a piece of paper, and a pencil). Generally, students are given a choice of four answers from which to choose. Most of this kind of testing is based on memory and does little to assess actual communicative ability. A student could score well on such a test and still be unable to hold a basic conversation in the target language. It has been my experience that as soon as a more communicative approach is introduced into the curriculum, students quickly improve in basic communication: question formation, eliciting responses, and the transmission of information. A communicative approach is necessary because Japanese students do not seem to naturally transition between written test and oral simulated task.

CLT as well as Performance Language Testing attempt to test a student’s Communicative Language Ability (CLA). Bachman (1990) says that CLA “can be described as consisting of both knowledge, or competence, and the capacity for implementing, or executing that competence in appropriate, contextualized communicative language use” Bachman (1990). This is to say that CLA is a unified system consisting of many aspects of the testee’s ability.

— Discussion —

This idea gives rise to the question, “what is it that we are testing?” A performance test samples behavior in a single setting with no intention of generalizing beyond that setting. In CLT, samples of performance are evaluated, in certain specific contexts of use, created under particular test constraints, for what they can tell us about a candidate’s communicative capacity or language ability (Weir, 1990). How does performance in tasks such as answering a customer’s questions about the accommodations in a hotel reflect language capability or competence?

In order to answer this question, we must look at the key aspects to performing a task. To interact with a customer in the hotel task the testee would have to use certain abilities

such as listening skills, grammar functions, and some pragmatic understanding to accomplish this task effectively. The proper use of these skills could be applied to a variety of tasks. There could be a list of key aspects of performance to be assessed such as fluency, appropriateness, accuracy, pronunciation, grammar, and so on (Brindley, 1990). Certain criteria could be established to evaluate proper performance in any given task.

Criterion-Referenced Assessment (CRA) attempts to answer this call for some continuity. In 1990, Brindley described the ability to evaluate a student's performance in one task as a reflection of the student's overall proficiency. He sees this proficiency on three levels. First, proficiency is a continuum that ranges from no proficiency at all to "perfect" proficiency. Second, the *criterion* is defined as an external standard against which learner behavior is compared. And third, levels of proficiency (or achievement) are linked to specific tasks (Brindley, 1990).

When considering any kind of testing, one must look at the problems that arise in using that test and what can be done to overcome those problems. Defining what there is to test is part of this problem. According to Alderson (1981) one of the problems with communicative language tests is the problem of language in use: "it is infinitely variable, being different for different individuals at different points in time" Alderson (1981). The levels of abstraction can be high. Tests are for groups of people, not for individuals. Alderson goes on to say that levels of abstraction are likely to be higher rather than lower in these tests: but it was argued that if one abstracts far enough from the situation or task, one reaches grammar, which is what language learners will need whatever they are going to use the language for, and grammar is the level of language most amenable to systematic description. This is the linguistic competence that discrete tests test. Linguistic competence plays a large part in the overall communicative. It is surely part of the testing apparatus and it deserves some consideration in regard to overall evaluation. But, how much, is a concern of the language educator.

Validity

Another problem has to do with validity. Is what we are testing a valid example of the communicative process? What we are testing needs to reflect what is covered in the syllabus. Validity is understood to involve many different aspects of a test. At this point I would like to introduce two types of validity: "face validity" and "content validity." Face validity concerns the question: is what we are testing a valid example of the communicative process? We can only answer this question by observing how effectively the testee is able

to communicate using the second language in certain target situations. In the test involving a hotel clerk scenario, a student might be asked to answer questions and give information about hotel rooms, and the test giver plays the part of a customer and asks questions, as would a customer to a hotel clerk. This could be seen to have a certain amount of “face validity.” If this were done in a setting resembling a hotel and front desk, all the better.

“Content validity” asks that the language being used by both the *customer* and the *clerk* be an accurate representation of a typical or even atypical conversation as long as the grammar and vocabulary were covered during the learning phase. According to Bachman (1990 p. 307) the main problems with demonstrating content representativeness are related to the adequacy of the sample. For testing purposes, a set of grammar and vocabulary to be covered in the test can be acceptable to establish some degree of “content” validity. In the case of the hotel clerk, an amount of unfamiliar grammar and vocabulary could come up and it would be necessary to consider how the testee responds to unfamiliar terms as valid for evaluation.

The greatest problem facing CLT is to determine sample size. Do we really need to know how a student will perform in all given situations before we put them behind the front desk of a hotel? Teachers and test givers are always limited on time for testing. Moreover, we must have a limited criterion set for a given test and follow that regardless of what other factors may be involved.

“Real-life”

One of the goals of communicative language testing is to recreate what happens in real life. This attempt at “real-life” language use as referred to by Bachman (1990 p. 329) can pose a problem for the language test. We need to produce a test that will create a real life situation in testing. This is referred to as the *authenticity* of the test. According to Bachman (1990 p. 303), there are three interrelated tenets that characterize the Real Life (RL) approach: (1) a view of language ability, or proficiency, as pragmatic ascription; (2) the reference to ‘real-life performance’ as a criterion, and (3) the belief that ‘face validity’ content relevance, and predictive utility are sufficient to justify test use. By reviewing these tenets in regard to a given test, we should be able to establish some kind of *authenticity*.

The first tenet, proficiency, exemplifies the need to show that by taking a given test the testee can demonstrate an ability to do X. If the testee is attempting to demonstrate his or her ability to answer a customer's questions about accommodations at a hotel, the testee would need to do just that, as defined by the criteria established beforehand. Explaining

about the rooms and answering questions about the services would be a reasonable part of this test. If a testee were able to do this with some degree of competence, then the testee would be proficient at this task. Working as a clerk in a hotel would also involve a whole degree of unpredictable questions and situations. One would hope that by demonstrating a degree of knowledge about the hotel and its services which would involve a certain amount of vocabulary and the grammatical functions to express it, it would not be unimaginable to assume that the testee is capable to perform in real-life. In fact, the testee might even perform better in real life due to the pressures involve in test taking.

The second tenet, the reference to ‘real-life’ as a criterion, would be met by the way in which the testing situation is designed. I work at one university that has a simulated front desk from a hotel with the name “University Hotel” on the wall behind the desk. Is it more *authentic* to perform the test there? Do we need to have an unknown person approach and start a hotel conversation following the criteria established for the test in order to fulfill the real-life criterion? I don’t think this is necessary or practical. The interaction between the testee and customer need only be imagined to be real. It may help the testee to get in character, but it is not a necessary component of the test. If the testee understands the situation, then the real life *criterion* is as good as it gets.

The third tenet is addressed in the teaching that is carried out before the test. Is the testee aware of the possible language and functions involved in the task? Video would help this phase with a variety of situations and exchanges being observed and demonstrated.

The Students

In conclusion, it is important to look again at who it is that we are testing. During my life here in Japan, I come across many different types of students. These students are studying a language for a variety of reasons. Some are studying in order to help them live and function in the society that uses the target language. This might be interpreted as the second language student. For these students, the tasks that they will need to become proficient are vast. Testing these students is secondary to learning the language properly and sufficiently. On a number of occasions I have observed students who score poorly on tests but are able to communicate and even create jokes within the target language. These students will go on with their lives and continue learning without tests. Many of these students are going to an English speaking country and that is the motivation and will be the *real-life* test for them.

Other students are studying the target language as a foreign language. These students

are doing this to get some kind of credit for gaining a certain amount of proficiency in the target language. This kind of proficiency can be reflected in test scores and evaluation. For these students, tests are the best way for them to gauge their progress and to see their weaknesses.

A third kind of student is not so interested in the target language. He or she is taking required classes in order to fill a requirement and may never get the opportunity to speak the language outside of the classroom. For these students, tests are a useless bother. All they want to do is pass the test. They will memorize the needed grammatical forms and repeat the sentences in order to pass. Once the test is finished, they will forget all they studied to pass the test.

Like all teachers, I am evaluating my students throughout the year and the final *test* is really just an exercise in effort and performance. I am trying to give my students a good experience with English, one that will encourage them to continue studying. Testing is a very important part of teaching, but it is only one part of the curriculum. Using CLT we can design tests that can make testing more practical, applicable, and a tool students will recognize and embrace.

REFERENCES:

1. Alderson, J.C. (1981) *Report of the discussion on communicative language testing* ELT Documents 111. London: The British Council
2. Bachman, L (1990) *Fundamental Considerations in Language Testing* Ch 4 and Ch 8 Oxford: Oxford University Press
3. Brindley, G (1991) *Current Developments in Language Testing* Singapore: Regional Language Center
4. Oller, J.W. (1979) *Language Tests at school: A pragmatic approach*. London: Longman
5. Spolsky, B. (1985) *Language Testing, 2.2* Bar-Ilan University
6. Standard Speaking Test: Manual (2000) ACTFL-ALC PRESS
7. Weir, C.J. (1990), *Communicative Language Testing*. London: Prentice Hall