

Going Beyond Parroting: Testing 4 Competencies Using Criterion Referenced Testing in Oral English Classes

Todd Armstrong

Abstract

The following paper presents how by carefully constructing criterion-referenced tests and helping students prepare for the test the 4 basic language competencies can still be achieved by learners in introductory oral English classes. The paper outlines what those competencies are and shows how to test for them.

There are two versions of the 3Rs. People of a certain generation will think of reading, writing and arithmetic. While people of a different generation will think of read, relate and regurgitate. The latter image is especially connected to tests in which many learners put a great deal of strain on their short term memory, commonly known as cramming, answer test questions during the designated test taking time but are unable to answer the same questions two days later. So instead of taking a snapshot in time as to the amount of material that was learned in a semester for example, perhaps there are ways to measure the basic competencies of language along with what class material was learned. That is what will be presented and discussed in this paper.

The paper will proceed as follows; a definition of the two common types of testing in language learning; norm and criterion referenced will be explained, next what the 4 basic competencies in language are and an example of using an interview test to test these competencies along with a few consequences of using criterion referenced interview tests for teaching. Some criticisms of using this type of testing and the consequences for teaching will be dealt with. Finally, a course can be viewed as to prepare students for an exam. There are some famous criticisms of teaching to the test and not promoting real learning which will be addressed towards the end of the paper but given the structure of

the exam and the priority placed on 3 of the competencies versus the competency that most learners have focused on in their formal language learning experience, this style of preparation can help learners move away from parroting and the modern three Rs. Before outlining the two types of testing, let us define what exactly is parroting.

Parroting is defined in this paper as mimicking and being able to verbally reproduce even the most complex syntactical and grammatical sentences by rote without ever truly ever understanding let alone being able to translate back into the interviewee's first language. An example is that many people are able to sing deep and meaningful songs in an L2 with perfect intonation, pitch and pronunciation but cannot have a basic conversation with someone in L2. The test taker can simply memorize certain responses and then on cue give the correct response. Questions such as giving memorized responses to questions based on pictures, such as "What is the boy holding in his hand?" "A book" would be the answer uttered. These rudimentary answers are fine and good but they do not measure the more communicative competencies which is what should be measured. Let us define what those competencies are more concretely and examine each competency in more detail to get a true understand of each.

Canale (1983) define communicative competence as consisting of grammatical competence, strategic competence, sociolinguistic competence and finally discourse competence. These are the four competencies that should be assessed when any type of interview exam which can be assessed using a criterion referenced interview exam.

Grammatical competence is concerned with knowledge of the language system. The system is simply defined as being the words of a language and how the words are sequenced to produce meaning. The competence implies that that this knowledge be of sufficient depth to accurately understand and express meaning. This involves sufficient knowledge of spelling, pronunciation, word formation, sentence formation and how pronunciation weakens or reduces in connected streams of discourse and so forth.

Sociolinguistic competence is concerned with language appropriacy in a given social context. Language is a social construct and thus carries with it conventions of social appropriacy. An example would be a server in a restaurant with an unfamiliar customer would say something like "Would care for anything to drink?" while the same server would utter the phrase to a more regular customer "The usual?" to convey the same message.

Strategic competence describes how we use strategies to compensate for linguistic deficiencies or enhance out communicative message. Second language speakers are often tasked with doing something they do not have

the linguistic resources to do. They need to compensate for their linguistic deficits by using some sort of strategy, for example talking around an unknown word, asking for help, using gestures, giving examples and so on. Strategies describe a variety of linguistic devices such as:

- *hesitation devices
- *repeating key words
- *shadowing
- *asking for repetition
- *paraphrasing
- *backchanneling
- *clarification questions
- *involvement questions
- *circumlocution

Discourse competence concerns one's ability to use English in different genres, for example: oral narratives, a scientific report, a business letter, computer manuals, text messaging and so forth. Each discourse type comes with its own set of rules and conventions. To function properly within these discourse fields requires an understanding of how these strings of language are joined together. Given that we have a better understanding of what the four competencies are being assessed, let us now try to understand what the difference between norm and criterion reference testing is. Both tests will be explained in their purest of forms, leading to the realization that especially for false beginners, a criterion referenced test is highly recommended with the most important elements of the norm referenced added.

McNamara (2000) outlines best the two distinctions "in terms of test purpose is that between achievement and proficiency tests." These tests are associated with the process of instruction. This process of instruction will be elaborated upon later in the paper. Examples are end of course tests, portfolio assessments and observational procedures for recording progress on classroom work and participation. In other words, achievement tests measure what the learner has learned by the end of the course. Criterion referenced tests are used in many instances which maybe more familiar to readers such as a drivers's license written and road tests. A learner takes a driver's manual, and attempts to display that knowledge by answering questions on a written test and a road test. There is a fixed amount of knowledge that the learner has to display to receive a driver's license. An important element is that the test taker's display of knowledge is compared to that which some governing body has deemed necessary to operate a motor vehicle safely. Contrast all of that with a norm referenced test.

Brown (1989) defines a norm-referenced test as:

Any test that is primarily designed to disperse the performances of students in a normal distribution based on their general abilities, or proficiencies, for purposes of categorizing the students into levels or comparing students' performances of the others who formed the normative group.

The two elements of norm referenced tests are the producing of rank order, comparing one test taker to another and the name norm referenced itself, referring to a normal distribution. Let us deal with these two ideas briefly to see exactly why these two points alone make the case for criterion referenced interview tests.

One is to produce rank order meaning that the purpose of the test is to produce a distribution around a mean. For example, the average score of a 50 test item test is 30 while some scored between 30 and 40 and even fewer scored higher than 41. The same can be said for those who achieved a score between 20 and 30 while even fewer scored less than 19 to zero. If you were to plot these results you would get a standard bell curve or a normal distribution. Some as Davies (1989) make the point that a valid test would produce such results. As we have seen with driver's licenses not mention also any other type of licensing examinations are more than valid and used in all areas of life. So, the idea that rank order must be produced for a test to be valid is not true. Also an assumption is made by those who favor norm referenced tests.

The underling assumption is that the class is normally distributed to begin with which is hardly the case with introductory conversation classes. Program requirements, various prior experiences with learning English, different motivational levels plus numerous other factors contribute to evidence that a non-normal distribution is mostly present in most classes. The learners are more the same then they are different which defies the definition of a distribution. Therefore, it is theoretically possible and legitimate to have all of the class pass the test due to the fact that most students were at the same starting point to begin with so it is more than plausible to have everyone past the test if the bar is set low enough. However, that criticism can be said for any test which therefore does not make criterion reference illegitimate. So how can a test which has the potential to be very easy, to skew teaching results be made more difficult?

The element of randomness, much like in many verbal interactions, is included into the test. Not complete randomness however, only questions that the material was done in class. Randomness in the sense that during the interview exam, two students interview each other while there is monitor that records the answers and then later makes the assessment of the scoring result based on pronunciation of both the questioner and responder and the four basic competencies outlined earlier. The questions are cards of 6 by 4 inches with the questions that were highlighted in class. The randomness is that of a basic shuffling of the cards so that each pair do not know which questions they will receive. Also, the next pair will not know what questions they will receive either, so that the element of

collusion, once one pair is finished they relay what questions were asked on the test, is greatly if not entirely eliminated. Therefore, highly competent people can still achieve a high score even they miss a few classes. However, most learners are low in their communicative competence plus the four competencies that coming to class is imperative.

It is in class where the students are told exactly what types of questions will be given. The exact same questions, word for word, will be presented in class and that students are encouraged to take notes and write the questions down. The questions in lay man's terms are open ended, meaning that there is a range of answers of acceptable answers, but given that it is the 4 competencies that being tested parroting answers demonstrates the limited competency the test candidate has. Possible answers are practiced in class. A detailed example will be presented next, discussing food in Japan.

An example of an open ended question for the oral exam is "Can you describe in 4 sentences a dish that is suitable for a vegetarian in Japan?" In class, a fixed format of given a basic list of ingredients, how they are combined by the various cooking styles, what the dish is served with and whether the interviewee themselves like this dish or not would be expected. All of these would be presented in class, practiced but it is up to the interviewee to combine these structures in a grammatical, sociolinguistic, strategic and discourse competent manner. How the interviewee can demonstrate these four competencies will be explained.

The obvious grammatical and formal competence; first does the interviewee actually understand question but if not there is a case for strategic competence to be displayed as to hear of keys words in the sentence such describe, suitable and vegetarian. So even though the interviewee might not be able to repeat the same question verbatim, given that the interview can identify key words, a form of strategic competence, the interviewee can still respond to the question. Now evaluation of answers again really shows whether the learner has the four competencies or not.

Obviously a dish that would contain any animal products would be viewed as incorrect answer even though it could be presented grammatically perfect. The sociolinguistic competence would be shown as to how it would be explained to monitor, the strategic competence can be demonstrated through gestures such as trying to explain skewers. This word has particular problems for many false beginner learners but if the test taker were to gesture somehow such as putting food on a stick, then the monitor, who should calibrated such as they are in Eiken STEP examinations, would be more than justified and legitimate in rewarding

scores for such attempts to communicate. Finally, linking words such as first, then and other linking words which are commonly found in a recipe allow the test taker to show discourse competence. This is just one of the open ended randomly generated within a fixed criterion that can really allow a test taker to demonstrate the four competencies and the knowledge learned in an introductory conversation course.

The purpose of this paper was to prevent an argument for using criterion referenced exams to test the 4 general competencies of learners in introductory oral English classes. Contrasting the difference between norm referenced and criterion referenced leads to a strong conclusion as to how the grammatical, sociolinguistic, discourse and strategic competency of test takers to be demonstrated by test takers. Norm referenced testing is still widely used because of discreet testing items that can placed on a multiple test taking sheet and marked efficient and at low cost by a computer. However, there has been substantial progress made and as exhibited by the TOEFL oral speaking section that can lead to same efficiencies and cost performance that can allow oral interview tests to be marked by computers. When that can be done, then truly really oral competency assessment can be the target at all levels of English education around the world.

Works Cited

- Brown, James Dean and Thom Hudson (1989) in Criterion Referenced Language Testing, London:Cambridge University Press.
- Canale, M. (1983) in J.C. Richards and R.W. Schmidt (eds.) Language and Communication, London: Longman.
- Davies, Alan. (1989) in Language and Education Performance on ESL Examinations: Is There a Language Distance Effect Volume 12, Issue 1 March.
- Glasner,R. (1963) in American Psychologist, New York Vol 136 Issue 1 March.
- McNamara, T. (2000) in Language Testing, London: Oxford University Press.