外国語として英語教育のオーラルコミュニケーション試験作成

Creation of speaking tests for EFL communication classes

Creation of Speaking Tests for EFL Communication Classes

Eric HAGLEY

外国語として英語教育の

オーラルコミュニケーション試験作成

ハグリー　エリック

Abstract：When teaching a communication class, any achievement test that wishes to have construct and content validity should include communication tasks. As there is emphasis on oral communication in the syllabi of most communication classes, again, it should be imperative to include oral communication tasks in any test associated with said classes. Doing so will also ensure positive backwash from the test is attained. In this paper the major differences between oral communication language testing and other forms of language testing will be discussed and a method of oral testing proposed.　The pitfalls and benefits associated with oral communication testing are also covered.

Key words：oral communication, English as a foreign language, testing

1. Introduction

Butt et al. (2000, p 5.) look to define language in terms of its functions.　They take Halliday's (1994) view that language has developed "to talk about what is happening, what will happen, and what has happened (the ideational metafunction); to interact and/or express a point of view (the interpersonal metafunction) and to turn the output of (these) into a coherent whole" (the textual metafunction). From this we can assume that the primary goal of language is to articulate in regard "happenings", "interact" and "express" oneself while being "coherent".　These points are the essence of what communication is. Any test of a student's language acquisition achievement should therefore also include a test of their ability to communicate.　It follows that any test should involve the student interacting and expressing themselves with semblances of coherence at the very least. This should be the essence of testing language and communication ability particularly so in regard oral communication.

At present, the vast majority of exams in communication classrooms center around listening, vocabulary and structure. Discreet point testing of items in these categories can show a student's understanding of them. Numerous internationally recognized tests incorporate these methods. However, the essential element of turning the output of language learning into a "coherent whole" is not able to be tested in this way. The aim of most communication teachers is to ensure their charges can communicate in the language they are teaching. Yet their exams do not reflect this goal.

Types of tests

Hughes (1989, pp 9-20) outlines the variety of language tests available to language instructors. Many of these can be incorporated into communicative language testing but they can not encompass all that is required to attain a valid and reliable understanding of a learner's language capacity, that is the ability a student has to produce language.   The most well known test type is the proficiency test. These are various in nature but have at their basis a desire to "measure people's ability in a language regardless of any training they may have had" (ibid, p 9).   If they are for specific purposes they can test the understanding of language used in a variety of situations that suits the stake holder who wishes to know whether the learner can function linguistically in a specified field.   Problems arise when trying to determine what constitutes the particular field to be tested.   This is amusingly pointed out by Bachman (1990, p 312) where he tries to determine what would be required linguistically of a New York taxi driver.   The infinite variety of options that are available to a language user involved in a speech act ensures that one will never be able to test all the situations that could theoretically arise, but a basic appreciation can be attained if sufficient aspects are covered to the satisfaction of the relevant stake holders.   This however will not tell the stake holder nor the learner of the learner's overall language capacity.   A more generalized proficiency test aims to do this, but as Hughes (p 10) points out the test "should have detailed specifications saying what it is that successful candidates will have demonstrated that they can do" not just give a figure that has no specific meaning behind it.

Achievement tests are those used by teachers and others to check the progress and what has been attained by students during and at the end of a particular course.   They can be highly restrictive in that they measure what a student has done throughout the syllabus and check the student's understanding of the material used in the course.   If the teacher bases the test "on the objectives of the course" (ibid. p 11) then a more flexible approach can be taken, as course objectives generally include the synthesis of all materials and exercises done throughout the course.

Another test that is highly restrictive, as it is for a specific purpose, is the placement test.   Required to attempt to put learners in a class where they don't feel out of depth, these types of tests can be

particularly difficult for students as they are often given to students who have no confidence in the second language, the test only reinforcing their fear of the language.　These tests should ideally be given in an unthreatening way however this is easier said than done. In order to be more specific in regard exactly what language knowledge has been attained and what is yet to be understood one can draw on diagnostic tests.　With the results of these, teachers can build specific programs to assist students in the areas that they are weak.　The analysis again is limited by the length and depth of the test.　They are powerful tools but like all the best tools they are difficult to produce and find. The placement and diagnostic tests are specific and do not lend themselves to testing a student's communicative ability.

Unless a test tests what Weir (1990) refers to as "the capacity to use language communicatively thus involves both competence and demonstration of the ability to *use* (my emphasis) this competence" it is laking in regard its propensity to gauge a student's communication ability.　Tests that allow students to use the language they are studying encompass both the language competence *and* performance that Chomsky (1965) refers to along with the ideas of language competence and "skill in using (language) knowledge" that Canale and Swain (1980), and Canale (1983) refer to.

Creating a test for oral communication classes

Firstly an understanding of the difference between direct and indirect testing is required.　A direct test encourages the learner to use what they have learnt in a practical way.　For learners of a language this is generally what they want.　For this reason, the direct test has the added benefit of encouraging learners to "practice the skills we wish to foster" (Hughes, p 15) and can be called a "helpful backwash effect."　Indirect testing attempts to test the underlying abilities that are part of language knowledge (Bachman, 1991) and competence (Chomsky, 1965).　When discrete point testing (testing one element at a time) is used in an indirect approach, very specific test scores can be attained.　However as there is no synthesis or evaluation involved in doing this type of testing it lacks the higher levels of thought that educators such as Bloom (1956) promote.　Integrative testing promotes these.

The results of these types of tests can then be outlined in a number of ways.　If the learner is compared in some way to other learners and then placed in a bracket that is tied to the results of the other students, this type of test becomes a norm referenced one.　If however the learners' results are explained with a description of what the learner was able to do the test can be said to be criterion referenced.　Again this form of testing, by giving information on what the student can and can't do, can become a powerful tool to empower the learner.　The backwash of such exams, particularly oral

communication ones, to any in-class speaking exercises can also be pronounced (Shohamy, et al. 1986).

To create a communicative test of language that truly measures the language capacity of the learner - *both* what knowledge is contained *and* what is able to be performed - a combination of the above approaches is required.   Both knowledge and the ability to apply it are required for someone to be considered truly competent in a language.   With only one or the other, the learner will come across as being in some way incapable when interacting with people of the language and culture being studied.   Both need to be taught and if testing is, as it should be, used to motivate the learner to learn more, both need to be tested too to let the learner know where he/she is on the learning curve and what it is they need to continue to work on.   This is where communicative testing comes to the fore, as it has the scope to take on this task, whereas the other forms outlined to date are usually too narrow in their appraisal of the learners' language capacity.

The most complete assessment of what it is that needs to be tested to gauge a learner's capacity is outlined by Bachman (1990).   The competencies he outlines make up the content areas of a total language capacity.   Many problems have been associated with trying to test so many aspects in the model, and indeed with the concept of communicative testing as a whole.   These are not insurmountable.

Objectivity in testing is another important factor.   When a marker of a test must make decisions on whether an answer is correct, better than another, or has some other characteristic, then the test becomes a subjective one.   If there is only one "right" answer with zero variability the test is said to be objective.   This brings us to the problem of how to measure language and communication ability. By using numerical figures to gauge language capacity we are in affect using a different language to describe ability.   Without a detailed description of what the numbers mean they can not really tell us anything about the learner's understanding of and ability to use the parameters outlined in the models proposed.   Thus we come to a crossroads where any way you go you are going to strike paradoxes. Tests lose some of their tangibility if one cannot put a figure to an outcome, yet figures reduce the face validity of a language test.

If we use criterion referencing we are basing our criterions on what is considered normal for a particular level of criterion by our standards, hence we are also creating a norm referenced test, the norms being those that the language instructor sets.   Quite a conundrum!   Lantolf and Frawley (1988) are particularly scathing of the AEI guidelines and indeed criterion referencing as a whole. They seem to believe that there is a need for an "in-depth consideration of proficiency that is independent of the proficiency test itself." They feel that if "proficiency is what proficiency tests

measure" then that is not enough.　　Their approach to criterion referencing is very minimalistic.　　If one took this view with all things that are to be measured, nothing would ever have a "final" measure. Does one require an in-depth consideration of height to measure one's height?　　Is one's height not what a height measure (ruler) measures?　　When one is asked "how tall are you?" does one reply "192.000001 cms - not a micron more not a micron less, but I'll let you know if there is any change when I can find a more accurate measuring center. I hear there's one coming out that can measure to the billionth of a centimeter!"　　The concept of infinity is one that applies to the measurement of anything, and as Fromkin et al. (1996) point out, language too, is infinite.　　If one gets wrapped up in it one cannot stop.　　In there notes following the article, Lantolf and Frawley do note that the metric concept has its place already, as too the scalar measure of color.　　Their criticisms of the guidelines and criterion referencing seem tied to the admirable belief that there is a lack of social justice in any scale.　　They advocate banning it and other criterion referencing until further research is done.

If all considerations for the learners being taught are taken into account, as seems to be the case with the model proposed by Bachman and Palmer, and included in the criterion referencing, a complete model can be created that also has various calibrations.　　Just as one can measure someone by saying "She is tall" or "she is over 6 feet" or "she is 192.1 centimeters" depending on what information one wants, a variety of criterion scales can be developed that will satisfy the needs of each individual tester and testee.　　It is by means of these criterion, and having skilled, trained people who are knowledgeable of all facets of communication, that one will be able to measure a language user's proficiency reliably and validly.

The next problem is where, when and in what manner can a tester test a learner's language capacity yet still ensure that the language being used is "real".　　The very fact that it is a test means the student is in "test" mode, not natural communication mode.　　There is no practical solution to this problem.　　The only way one can ensure that the test is a real reflection of the language capacity of the learner is to observe the learner in a number of situations over a period of time.　　Another problem that arises with this type of communicative testing is how it is to be operationalized.　　If it is to retain total validity, it requires an oral component.　　Unlike a pen and paper test where the tester can take the paper away for careful and considered marking, spoken words disappear unless they are recorded.　　It should be standard practice for any oral test to be recorded in some form to ensure that careful and considered marking can take place by "multiple independent scorers" (Hughes, p 42) to ensure reliability.　　There are problems with recording conversations, as the majority of students feel very self conscious.　　This "self consciousness" can be reduced by ensuring that from the very outset of language instruction, recording is a part of every class and part of homework.　　Students become more relaxed with the concept early in the syllabus and are used to it by the time the exam is administered (Hagley, 2003).　　The first recordings can be used as a marker for any achievement

based testing that is to be done, and a portfolio of their work can be developed - a great motivational tools that can give excellent backwash as the students have a clear gauge of their progress.   If the criterion being used to adjudge the learner are explicit, and the instructor goes over with the student these criterion they can be used in any language counseling sessions with the student.

The actual tasks then become the pivotal point on which communicative testing rests.   Weir (1990, p 12) points out that "in assessing the ability to interact really we should try to reflect the interactive nature of normal spoken discourse and attempt to ensure that reciprocity is allowed for in the test tasks included". This includes the unpredictable nature of language, the ability to negotiate meaning and cope with other elements that make up strategic competence in the Bachman model.   Simple interviews between the tester and the testee cannot cover all of this.   A variety of interactions need to be included.   In oral communication classes, put emphasis on this is paramount, but other testing procedures should also be carried out to ensure an overall understanding of the language knowledge capacity is also attained.   To do this there is a place for discreet point testing and other forms of indirect testing, but these should not be the major component of any test of language capacity, rather they should be there to complement.

One example of a communicative test

As mentioned above, if recording of conversations has begun at the beginning of a syllabus and continued throughout students will be comfortable using the recording device by the time of the exam. Their recordings will also be a valuable tool for them to reflect on what they say and how they say it. Teachers will be able to give advice to the students based on those recordings. This in effect becomes the first part of the communicative test as the degree of progress attained, as shown in the portfolio of recordings, can be given some weight when giving final assessment.

The major component of the proposed test will be the spoken exam. The administration of this exam should be such that practicality, reliability and validity are maintained. To do so, the option chosen is a role play with two or three students participating in it. The roles given to the students need to be very specific and ensure the need to use the material that was taught throughout the syllabus. The role cards should be given to the students at least three weeks prior to the speaking exam taking place. This is so the students know precisely what they need to do. On exam day, students are assigned partners randomly as too roles. As this is the case, they cannot simply memorize a conversation. They can practice but they cannot memorize as improvisation with the random partner is required for communication to take place.

There are numerous possibilities in regard roles. Brindley (1991) suggested creating a bank of assessment tasks and this has occurred in more than one institution around the world. The majority of these have been for written assessment tasks, though there are speaking component examples too. Trialling of each task to verify reliability is an essential part of the process. Any speaking tasks that are created should follow this process too.

On exam day the teacher, after assigning partners and roles, gives the students time to set up and test their recording devices. This should not take long as they have used them throughout the semester. Students should then record their names, student numbers and which role they will have. Teachers can use this section of the recording to put a voice to the name of the student when playing back the recordings for assessment. After all students have finished doing this at the same time, the teacher says "begin the exam". Students all start their role plays at the same time. This ensures that the exam is practical in that time-wise it doesn't take a long time to administer. Teachers should give the students at least 10 minutes to finish their role plays thus giving them ample opportunity to show their communicative strengths. Obviously, depending on the content to be tested this length could be longer or shorter. Students are using the material they have learnt in a communicatively productive way. This ensures construct validity of the test is maintained.

As outlined above, final assessment of tasks completed, in this case the role play, can take a number of forms. Criterion referencing is one method. As the role plays are recorded reliability can be assured. Another, perhaps more controversial method, would be to take an average of the number of words spoken per minute from a random five minute sample of the ten minute role play. Noting the number of communication breakdowns and who was at fault in these could be another factor that affects final assessment. Ideally the assessment would also include a breakdown of the different areas each student is required to work on to improve with this information being returned to the student. This is one area where criterion referencing is particularly useful.

Conclusion

From the above discussion communicative language testing can be seen as a method of testing that has at its core the desire to gauge a learner's language capacity in such a way as to also provide the learner with beneficial feedback on what aspects of their quest for a greater language capacity need addressing. It does this by adopting the position that language testing should revolve around the interactive nature of communication.   Problems remain as to how to actually calibrate a measuring device, but work with criterion referencing based upon the model of language as outlined by Bachman and Palmer has gone a long way to assuaging these problems.   Problems regarding testing methods have also caused angst for those committed to this form of language testing but a variety of testing methods that

encourage interaction and an acceptance that testing should be over a period of time in a number of situations has ensured that these problems too have been overcome. The proposed oral test is one such method and should be extensively trialled. A question bank of speaking activities can also be developed.

Bibliography

Bachman, L. (1990).   *Fundamental Considerations in Language Testing.*   Oxford: Oxford University Press.

Bachman - Palmer. (1983a). As quoted in Bachman, L. 1990.   *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Bloom, B.S. (Ed.) (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. New York ; Toronto: Longmans, Green.

Brindley, G. (1991). "Defining language ability: the criteria for criteria" in S. Anivan (ed.) *Current developments in language testing.*   Singapore: Regional Language Centre.

Butt, D. et al, (2000). *Using Functional Grammar - An Explorer's Guide* NCELTR, Macquarie University.

Canal, M. (1983). *From communicative competence to communicative performance* in J.C. Richards and R. Schmidt (eds.) (1983) *Language and communication.*   London: Longman.

Canale, M. and Swain, M. (1980). "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing" in Applied Linguistics, 1, 1. pp 1-47.

Chomsky, N. (1965). *Aspects Of The Theory Of Syntax* Cambridge, Mass. MIT press.

Fromkin et al. (1996). *An Introduction To Language* Sydney, Harcourt Brace & Company.

Hagley, E. (2002). "Improving Oral Communication Pedagogy Through The Use of Tape Recorders" in JALT Hokkaido Journal, 6,

Halliday, M.A.K. (1994). *An Introduction To Functional Grammar* Oxford University Press.

Hughes, A. (1989).   *Testing for Language Teachers.*   Cambridge : Cambridge University Press.

Lantolf, J. P. and Frawley, W. (1988). "Proficiency: understanding the construct" in *Studies in second language acquisition*, 10, 2. pp 181-195.

Shohamy et al. (1986). "Introducing a new comprehensive test of oral proficiency" in ELT J.; 40: 212-220.

Weir, C. J. (1990). *Communicative Language Testing*. London: Prentice Hall.

執筆者紹介
氏名：ハグリー　エリック
所属：室蘭工業大学大学院工学研究科ひと文化系領域
Email：hagley@mmm.muroran-it.ac.jp