# The Importance of Oral Testing for Conversation Classes
## A Framework for Planning Oral Tests

David R. Evans[1]      John Herbert[2]

1 National College of Nursing, Japan ; 1-2-1 Umezono, Kiyose-shi, Tokyo, 〒 204-8575, Japan
2 Hiroshima Prefectural University
　evans@adm.ncn.ac.jp

【Abstract】 The aim of this paper is two-fold. Firstly the paper will argue why oral testing is the best way to test 'conversation' classes at university/college level and secondly it will propose a framework that teachers can use to organise their own exams. Speaking is a different skill to reading and writing and therefore to test the ability to speak in any other way is not a test of speaking. Speaking involves both listening and speaking in 'real time' and this can only be properly assessed in spoken form. Oral exams are not used for 'high stakes' test because of the element of subjectivity that is unavoidable in the assessment of the spoken language. However, within the college environment, in which teachers are able to continually assess their students, this is less of a problem. The advantages and disadvantages of both traditional testing (particularly multiple-choice) and oral testing are discussed. A framework for oral testing is outlined, discussing the key issues that a teacher needs to consider when organising oral tests. This includes a discussion of whether the test should be proficiency or achievement style, criterion referenced or norm referenced, holistic or analytical in grading, and also discussed some of the different kinds of formats that exist. A timeframe helping teachers to decide when to inform students is also given. It is recommended that teachers both give feedback to students individually in order to both encourage and direct student's study, and to solicit feedback from the students so that the test can be improved in later tests.

【Keywords】 English conversation，oral testing，testing methods，testing framework，testing feedback

## I. Introduction

Language testing can be such a daunting area, with its own body of technical terms, statistical analyses and the like, that many English teachers shy away from devising their own tests. Bachman and Palmer (1996) express the problems and reluctance with which language teachers face language testing.

> In virtually every group we have worked with, we have found misconceptions about the development and use of language tests, and unrealistic expectations about what language tests can do and what they should be like, that have prevented people from becoming competent in language testing. (p. 3)

The aim of this paper is to urge teachers who teach 'conversation' classes to use oral testing as a means of testing. There are many reasons why this should be done assuming certain practical considerations are fulfilled. This paper will：(a) outline these reasons ; (b) address the potential disadvantages of oral testing ; (c) explain why the advantages of oral testing outweigh the disadvantages ; and (d) provide a framework for teachers who wish to pursue oral testing with their own students.

## II. Reasons for implementing oral tests

Assuming that 'conversation' classes are concerned with improving students' conversational ability, then to test in any written form, be it multiple-choice or otherwise, can not hope to capture the skills that are needed for conversation. According to Cornbleet and Carter (2001) the following are the distinct characteristics of 'conversation', each of which acts as a reason for choosing oral tests over non-oral tests for testing conversation.

*Conversation takes place in real time*

There is a great difference between speaking and writing, in that in speaking we have to do it 'now'. This immediacy puts a great pressure on both our ability to

speak and to listen. In the written word, we are under less pressure as we have the chance to gather our thoughts and prepare the message we wish to convey, which is seldom the case in conversation.

Brazil (1995) asserts that the pressure of having to speak and listen in 'real time' makes these skills problematic for the native speaker of a language.

> Our ability to put together what we want to say may not always be equal to the pressure to keep up with ourselves, so to speak, in the delivery of our message. As listeners, too, we frequently feel ourselves under similar pressure. The fact that time is passing makes it imperative to decode what we hear promptly so as not to miss what comes next. (p. 11)

In speaking a foreign language this ability is even more difficult, and this very essential skill can not be tested unless the test is spoken. Another consequence of speaking in real-time is that because of the pressure of time we do not communicate in 'sentences' but in 'utterances'. David Brazil believed that "speech is characteristically used in pursuit of purpose…the practice of inventing a sentence…is a practice of the sentence grammarian, not the user". When we speak, we generally do so for a reason and 'getting our meaning over' is more important than being grammatically correct, so speech does not conform to the same rules as written English.

Another aspect of speaking in 'real time' is that we tend to use simpler, less complicated language. There are some exceptions, such as in a lecture, but this is because we have prepared what we are going to say beforehand and, consequently the normal limits to speech do not apply. In particular circumstances, such as if we need to make a difficult request to our employers, or perhaps if we are rehearsing for an argument that we intend to initiate, such a strategy might be employed. However, for most conversation, there is no prior rehearsal, which means that what we say is simpler and often 'grammatically incorrect'. Students need to be aware that all native speakers, from David Beckham to George Bush, make mistakes when speaking, and do not speak in 'whole sentences'. One consequence for the language student is that it is not necessary to form a whole sentence in their head before speaking, which often impinges on fluency. Many advanced non-native speakers of English can sound unnatural or stilted precisely because they are trying to speak 'written English', instead of concentrating on conveying their meaning, which is the function of the spoken word. At a more

general level, it is often the perception that English must be spoken perfectly or not at all that deters many students from actually trying to speak English. This fear of mistakes is constantly reinforced by written tests that often measure only accuracy. However, making mistakes is not in fact as important as traditional testing would have us believe.

*Conversation is face to face*

With the exception of telephone calls (and discounting online "chat" which takes place in the written medium), when we converse we are in the presence of the person to whom we are speaking. This means that we receive more information than we do when reading a text. We can see emotion and interest in the faces of those to whom we are speaking and adjust what we say accordingly. If we think our conversation partners look bored, we will change topic ; if we think that they look keen to speak we will break off and allow them to do so. We might change what we were going to say if we can see our conversation partner is upset, or angry for example, or if we think what we have said has been misunderstood. Facial expressions might make it unnecessary to complete an utterance. When we see that our point has been understood we do not feel the need to finish what we were saying. These are all aspects of conversation that can not be tested unless the test involves actual conversation.

*Conversation is interactive*

Conversation takes place with other people, so social skills are involved. One of the most important is that of 'turn-taking'. This means that the role of speaking will alternate between the participants of the conversation. Even if the conversation is being dominated by one individual, the listener still has an important role to play in oiling the discourse by nodding or shaking his/her head, or by displaying interest, boredom, sympathy etc. Knowing when to speak and when to give an opening to a listener to speak is a skill that students need to learn and practice, and the ability to do so in a foreign language, can only be tested orally. In group conversations we will be aware of those who have not had a chance to speak and we might ask questions to involve others. This is another important skill that can only be tested in a group environment.

Perhaps the most important reason for testing conversation with an oral examination, is because conversation involves both listening to what someone else has said

and responding to that utterance, something that a written or listening exam can not do in the same manner.(For a detailed discussion of assessing listening, see Buck(2001)and for assessing reading see Alderson (2000)). Such a test may be able to measure receptive skills, but this is not what conversation is about. Heaton (1998)explains the problem as follows,

> In many tests of oral production it is neither possible nor desirable to separate the speaking skills from the listening skills. Clearly in normal speech situations the two skills are interdependent. It is impossible to hold any meaningful conversation without understanding what is being said and without making oneself understood at the same time.(p. 88)

## III. Potential disadvantages of oral testing

Having stated why oral testing is desirable for conversation based classes, it is important to look at the reasons why it is often passed over in favour of written tests of speaking. The two main problems usually associated with oral testing are concerned with reliability and practicality. The reliability problem according to Heaton(1988)is what he calls "mark/re-mark reliability". In high-stakes tests, such as university entrance exams, when there could conceivably be tens of thousands of test takers, it is impossible to be sure that students would get the same grade regardless of who is doing the assessing. Even with training, it would be impossible to guarantee that all assessors would award the same grade to the same candidate. Lazarton(2002) asserts that the grader "is not a neutral factor and must be accounted for in test validation"(p. 173). Anyone who has tried to grade orally will know that even when being the sole grader, it is hard to be sure that one's own grading is consistent. Therefore, in comparison with multiple-choice grading, in which all papers will be graded with 100% reliability, oral testing is undeniably less dependable.

This is not to say that multiple-choice, which is the preferred means of testing in most exams, is without flaws. In multiple-choice, there is a one in four chance on any question that a student can guess an answer. The extent to which guessing influences the final score is unknowable, but it cannot be denied that each student's final score will be higher or lower, depending on if they have 'guessed' the correct answer. In the words of Hughes,(1989), "the trouble is we can never know what part of any particular individual's score has come about through guessing"(p. 60). It is conceivable that this element of uncertainty in a candidate's score might be no greater than the error that might exist in oral testing due to variability in graders' ratings, yet this inconsistency is regarded as acceptable in multiple-choice and unacceptable in oral testing.

The second often cited problem of oral testing is practicality. For large-scale testing, it is far harder to organise oral testing because of problems with finding suitable assessors, the cost of paying them, and the logistics of getting the candidates and assessors to the right place at the same time, than it is to hold a written exam. Within the classroom environment these difficulties might exist if there are large classes and too many students, but for many colleges, these difficulties are not so acute.

## IV. Why the advantages of oral testing outweigh the drawbacks

The key problem with oral testing is that they are, as McNamara(2000)states, "necessarily subjective"(p. 37). One person, the rater, is judging another, the student, and the grade awarded is overall impression. No matter how professionally this is done, it can never be ruled out that there is error in the awarding of a grade. With multiple-choice, we can be sure that there is no bias or inaccuracy with the grading. There are, however, other problems. For a detailed analysis of its shortcomings see Hughes(1989, pp. 60-62), but to summarize, the key points are :

1 The technique tests only recognition knowledge.
2 Guessing may have a considerable but unknowable effect on test scores.
3 The technique severely restricts what can be tested
4 It is very difficult to write successful items
5 Backwash may be harmful
6 Cheating may be facilitated

All of these criticisms need to be considered before deciding whether or not to use multiple-choice. In the case of oral testing, the fact that multiple-choice can only test recognition shows that it is not suitable for measuring the ability to communicate.

Hughes' point that 'backwash' might be harmful is equally important. Backwash is the impact that exams have upon how subjects are taught. Equally, it must also have an influence on how students study. As a multiple

choice test is concerned with recognition skills and not use, teachers and students will understandably devote their energies to maximising the ability to recognise the patterns in English language that are frequently tested in multiple-choice. This does not mean, however, that these patterns can be reproduced actively. When testing is done orally, the positive effect of backwash will be that both students and teachers will concentrate on 'language use' rather than accruing what is sometimes referred to as 'knowledge about language'. This should also be beneficial in monolingual classrooms, where understandably students can be reticent to talk to their classmates in English, rather than their mother-tongue. When students realise that speaking English in class will help improve their test performance, the purpose of speaking becomes much clearer. Additionally, it should also concentrate the teacher's mind; as if the teacher decides to test orally this will impact on his/her teaching style. Once a commitment is made to oral testing, the focus of the class will shift further to practicing the skills needed for speaking.

In the context of the college teacher, the arguments for doing oral testing would seem to outweigh the disadvantages. For 'high-stakes' testing, from the viewpoint of practicality, it might not be feasible, but it would seem that the current status quo is not beyond approach.

Heaton notes that,

> Continuous assessment by the teacher, with all his or her classroom experience, is generally (but not always) a reliable method of measuring the oral skills. (P 104)

Oral testing in conjunction with continual assessment will be fairer than simply having one end of year exam to determine a student's grade. In all exams, some students will suffer badly from nerves, and an oral test will probably be hardest for those who are shy. However, as speaking in a foreign language does necessitate overcoming such fears, oral testing, with a teacher known to the student, should minimize this concern.

## V. Framework for oral tests

*How should you test? Choosing a 'direct' or 'indirect' test*

There is a consensus among testing experts that the best way to test, where and when it is feasible, is through 'direct' tests. Bailey (1998) defines direct tests as "those in which the learner's response involves actually doing the skills being assessed" (p. 75). Direct testing is possible for the productive skills of speaking and writing. For the receptive skills of listening and reading 'indirect' testing is used. As it is not possible to know if a student has read and understood an article, for example, questions may be written to tease out how much has been comprehended. Bailey points out that another weakness of this kind of test, is whether it does actually measure what it purports to do. She gives the example of the tenuous relationship between multiple-choice and reading, but the same is equally true for multiple choice and speaking.

> Someone who is good at selecting the correct response on multiple choice grammar items might not be an effective writer. (p. 75)

Bailey continues to say that indirect tests such as multiple-choice can have a negative effect on the way that students study the language.

> Another problem with indirect tests is that they may result in negative washback. For instance, if learners spend time studying bits of decontextualized grammar in preparation for an indirect test of writing, they may spend less time actually writing in the target language. (p. 75)

As mentioned earlier, this style of testing predominates in high-stakes testing, as it is easier to carry out and grade. Hughes (1989) believes that teachers should test the abilities they wish to develop so that "if you want to encourage oral ability you should test oral ability" (p. 49). If we test indirectly, Heaton (1988) asserts that, "we are removing an incentive for students to practice in the way we want them" (p 45). The important point here is that if students know they are going to receive a written paper to test their oral ability in English, then there is little incentive to practice in class. Instead of seeing English as a tool and a practical skill it will remain only as an academic subject, something to know about but not something to use. Having an oral exam should make the importance of practicing English explicit, and, assuming that students do actually improve, will be more motivating as they can see that practicing does lead to both an increased ability to understand spoken English and to speak it. Shohamy (2001) is another who is critical of the over reliance on the 'psychometric' testing tradition, in which the sanctity of the test is placed above all other concerns, concluding that.

> ...traditional testing is not interested in the motives for

introducing tests, in the intentions and rationale for using tests or in the examination of whether intentions were fulfilled. It is not interested in the steps taken in preparation for tests or in how test takers feel about tests and their effects on those who failed or succeeded in them. It also overlooks how the test affected knowledge learning patterns and habits.

Traditional testing views tests as isolated events, detached from people, society, motives, intentions, uses, impacts, effects and consequences. (p. 4)

Another of the growing band of malcontents in testing is Underhill (1987), who like Shohamy, believes that it is the student and not the test that should have priority in the testing procedure. His comments are concerned with oral testing.

"...oral tests must be designed around the people that are going to be involved. This is a human approach ; we want to encourage people to talk to each other as naturally as possible. The people, not the test instrument, are our first concern." (p. 4)

"...oral tests, because they involve a subjective judgement by one person of another, are likely to be less reliable ; but it suggests that the human aspect of that judgement is what makes them valuable and desirable. When we test a person's ability to perform in a foreign language, we want to know how well they can communicate with other people, not with an artificially-constructed object called a language test." (p. 5)

As the doubts grow about the validity of traditional testing and as greater attention gets given to the significance that the style of testing has in terms of washback, it seems that whenever it is physically practical, a "human approach" using direct oral testing will be a better choice of test for conversation than any other.

*Proficiency or achievement*

Once the important decision has been made to test 'directly' with an oral test as a test English conversation, the next decision to make is whether to choose a 'proficiency' style test or an 'achievement' style. The former is a general measure of a student's level of English, whereas the latter is a test directly related to the course content. An advantage of the proficiency style test is that it can have a wider scope and can result in more natural dialogue ; the downside is that it will not necessarily reward those who attended your class. Students, who were good speakers of English prior to joining your class, could in theory not attend any classes, but still come out with a very high grade. The possible danger of 'achievement' tests is that in an effort to use phrases or vocabulary learnt in class, natural-style usage will be neglected in an effort to display the fact that a particu-

lar phrase or structure has been memorised. The advantage is not just that students will be able to see the advantages to attending class, but that someone who previously had a low level of English, should, if the student has studied hard, be able to take the test on a similar footing to a student who had better language skills prior to the start of the course. Cohen (1994) believes that tests should be achievement based and directly related to the goals of the course and makes the important point that without doing this there is little incentive for students to participate in classes.

One of the primary reasons for conducting language assessment in the classroom is to promote meaningful involvement of students with material that is central to the teaching objectives of a given course. For this meaningful involvement to take place, the goals of the assessment tasks need to reflect the goals of the course, and these goals need to be made clear to students. (p. 13)

His final point, that the goals need to be made clear to the students, is also extremely important, as without this, they will be lost and unsure of what is expected from them. The goals should be made clear to the students at the beginning of the course.

*At the beginning of your course*

Ideally, a teacher should tell the students what is expected of them at the beginning of the course. This should include information regarding the importance of attendance, class participation, effort, continual-assessment or whatever else the teacher (and college) deem to be important, in terms of the final mark/grade given at the end of the course. If a teacher does not make it explicit that attendance will contribute towards the final grade at the start of the course, it would be unfair to introduce this as a factor just before the final exam. If a teacher is planning to use an oral test, this too should be made very clear, as if this is new for students it will take some getting used to. Students need to understand that the best means of preparation for such a test is to talk in English as much as possible, at every opportunity that presents itself, and especially so in class, as if it is an achievement test, any of the things done in class might be in the examination. In the first year of its introduction, it might also be helpful for students (and the teacher) to have a mid-year 'dummy run', so that any glitches can be ironed out and so that the format is not a total surprise to students. Of equal importance is that it makes it real to students that speaking is of primary importance in the course.

Students should be provided with information before the exam on the way in which they are to be tested, and how they will be graded. Without this information, they cannot usefully prepare for the test. If the aim of your test is to replicate real conversation, then it is important that students are not unduly anxious. It should be your stated aim that students are as relaxed as possible when taking the exam. If they are overly nervous their performance will be impaired. If you are able to do so, providing a 'practice test week' prior to the exam will help allay students' fears, especially if this is the students' first exposure to this kind of exam. If students can be persuaded to see the test as non-threatening, and if their grade is a combination of the oral test and continuous assessment (based upon participation) then the fear factor should hopefully subside. The first time students do an oral test, there are bound to be nerves, but once they are familiar with it and know what is expected they should overcome this.

*Choice of grading standard : Criterion-referenced versus norm-referenced testing*

When grading students there are two kinds of ways of doing it, using 'criterion-referenced' or 'norm-referenced' testing. Criterion referenced testing is when certain goals are set for students to achieve, and if they do so, they pass. Under this system if all students achieve the goal, everyone passes. Norm-referenced, on the other hand, is when students grades are compared with each other, and a percentage are awarded a pass (or A grade etc). Under this system it would be possible to have a class of brilliant scholars but only a certain percentage would be allowed to succeed. Clearly, criterion referenced testing is more suitable for college testing, but each college will have its own expectations regarding students' grades and this might take precedence. If the institution does give flexibility and allows criterion-referenced tests, then a generous marking policy can be beneficial. Research on motivation has shown that if students get good grades, their attitude towards the subject changes and their performance improves as a consequence.

*Holistic versus analytic grading*

The teacher then needs to decide whether to use 'holistic' (sometimes called 'impressionistic') or 'analytic' grading. Holistic is when an overall score is given for the student's performance, whereas analytic is when the student is graded in different categories, such as comprehension, fluency, complexity of English. The holistic style is easier and quicker, but less helpful when giving feedback, as there is no differentiation between the aspects that make up conversation. If a student is quite fluent, but also inaccurate, the scale is limited in differentiating between such differences and the rater must value one aspect over another. When using analytic scales, this problem can be avoided as students would get scores for both categories, but there is a danger that by concentrating on too many aspects that the overall impression of a student's communicative skills is obscured. If the scale has a limited number of categories and does not attempt to measure too much, these divisions should be useful for the students when they receive their feedback. In the college environment the teacher can choose analytic scales according to the aspects of conversation that most need development.

*Choosing the format of your oral exam*

There are many different styles of oral exams, and some might be more suitable for your students than others. The choice will depend on what has been done in class, what the goals of the English classes are perceived to be, and on which test is most likely to be best at revealing student understanding of the course. Time pressure may dictate if it is possible to use a mix of styles, but a process of trial and error is probably the best way of finding out what is most suitable for any particular course.

For a full overview of the different kinds of oral tests available, read Underhill (1987, pp. 44-87). For reasons of space only three kinds will be discussed here, the interview test, group testing and role-playing.

(a) *Interview tests*

Interview tests are one method of testing, in which the teacher talks to students individually. The main criticism of this kind of test is precisely because it is an interview and not a 'conversation'. The critics say that the test-taker is in an unequal power relationship with the tester, and this detracts from the test. Perret says that,

> ...not only is the interviewer a stranger to the respondent, meaning that the social distance between the two is great, but the interviewer is also in a role of considerable power, depending on the importance of the ratings that the interviewee receives. (cited in Cohen, page 267)

This is obviously a danger in high-stakes testing but

should not be a drawback in testing within the college environment. Furthermore, once aware of this drawback it should be possible to impress on students the importance of taking an active rather than a passive role in the interview. One consequence is that the test-taker does not ask questions, or take the initiative in introducing topics of conversation, but simply reacts to the tester (see Kormos, 1999) The more relaxed the student is with the teacher, the better the performance of the student is likely to be.

The interview test does have certain advantages as well. The teacher can use his/her expertise to ensure that the student is tested on all the aspects that he/she feels are important. This can be lost in other forms of oral testing when the teacher is not 'in charge'. It is also far more likely that when the conversation partner is a teacher, the utterances of any students are far more likely to be comprehended, something that is not always the case when the conversation partner is a fellow student. Yet another advantage is that some students are intimidated by speaking English in front of their peers, but can excel away from peer pressure. The teacher will also be able to increase the difficulty level of questions, answers and topic, whereas students will understandably be reluctant to use expressions or ask questions to classmates that they might not be able to understand.

(b) *Pair or group testing*

This style of testing avoids the problem of 'social distance' as students are required to talk to fellow students. It can also free the teacher from the role of interviewer and allows the teacher to give full attention to assessment. (It need not be done this way of course. The teacher can participate, and this might be very valuable if the group is having difficulties.)

A possible problem is that students might, if they know their partner (s) before the test, work on a prepared answer, which invalidates the aim of creating spontaneous conversation. If the students decide the groups, then this risk is greatly increased. Random assignment to groups by the teacher is likely to reduce this risk, but equally it might increase the possibility of students in groups not being compatible. This might seem trivial, but we all know from our own experiences in our mother tongue that we are less likely to converse successfully with people that we do not like, or do not know well.

Prior to the exam it is obviously positive if students practice, -this is after all one of the key reasons for testing in that it makes students review what they have studied-, but although the students should be able to know in general terms the style of the test, if too much detail is known then rote-learning does become a risk.

Thought needs to be given to how the test is begun. For topic-based tests it might just be sufficient to nominate one student to initiate the conversation on a particular topic, or to allow one student to choose the topic from a selection. Another possibility might be to give students prompt cards with a topic written on each. Having a choice is fairer on students as there might be topics on which they have no interest in their own language, let alone in a foreign one. A wide selection of topics or distinct sections within topics is also necessary so that news of the exam does not leak to those waiting to take it (care needs to be taken so that the different questions are of a similar level of difficulty). If students can make an educated guess as to what topic they will have to answer, then scripted answers become a possibility.

(c) *Role-playing*

Many teachers might feel that this is too contrived for an exam, but it can be useful in testing certain aspects. For teachers who teach E. S. P. (English for Specific Purposes), this kind of test can be very useful. (For a detailed review of assessing languages for specific purposes, see Douglas (2000). It would be too much to expect of students if it has not been a regular feature of classes, but if it has, it is perhaps the only way that speaking skills for particular situations can be tested. In a nurses' college, for instance, it is hard to see how else one can test a student's capacity to, for example, take a patient's blood pressure, unless it is done in a role-play format. The key point, as Bailey (1998) says, is that in any role-play, the situation should not only be one that the students' have experienced, but it should also be believable.

*Role of teacher during the test*

What the teacher does during the test will vary depending on the test format. However, for any of the three kinds of oral exam formats set out in section (7), it is important to have a short gap between each interview, group, or role-play so that the teacher/assessor can write down impressions, and/or provisional grades for each student. It is useful to record the tests, so that you

can review the tests afterwards if need be. If your testing lasts over a period of days you might want to refer to the tape before you recommence grading, so that you can see what kind of marks you were giving the day before. Recording students' performance also enables you to seek a second opinion if you feel the need to do so (taking into account inter-rater reliability complications) and, if you listen to it again, the tape may help you notice areas in which all the class needs help.

*After the test*

(*a*) *Feedback for students*

The more feedback you are able to give students the better. If students only receive a grade or a percentage, they are none the wiser about their strengths and weaknesses. Although it is time-consuming, adequate feedback can be an effective motivational tool. It also shows that the teacher is taking an interest in each student as an individual. As a minimum, students should receive a photocopy of your grading form. The more detailed information the students receive, the more helpful it will be. Comments on both the students' strengths and on the areas that need improvement should be included. If the feedback is only on student weaknesses, it could be demoralizing and therefore counter productive. As Cohen (1994) observes,

> Depending on the quality of the feedback and the attention the students give it, they may learn something about their areas of strengths and also about areas in which they are weak, prompting further learning or review. (p. 14)

If students receive a photocopy of your grading form, and then, having had an opportunity to digest what has been written, have a brief chat with the teacher (two or three minutes would be ample) the impact of the feedback will be heightened. Not only does this enable the students to question or clarify what has been written, but it gives the teacher an excellent opportunity to garner a host of further information from the student on such matters as how they are finding the course, what they find most difficult, why a certain student can speak so well, and so on.

(*b*) *Feedback from students*

Feedback from students is useful in enabling the teacher to improve the test, and, in incorporating the students' feedback in future tests, will increase the legitimacy of the test in the minds of the students. Student feedback can be gained from not only talking to students, but also in the form of a questionnaire. (Students can fill this in while the teacher is talking to students individually) It is highly unlikely that any teacher will hit upon a perfect format for a test the first time it is carried out, and if the written feedback is filled in anonymously and students are encouraged to respond honestly, the teacher is likely to glean very valuable feedback on the students' perceptions of the test. Such feedback may not just lead to changes in how the oral test is conducted in the future, but also might lead to changes in how the class is taught. For example, should students feel that they had insufficient vocabulary to cope, this area could be emphasised in future classes. When the teacher talks through the feedback with individual students, it is possible to find out why the students who performed best, did so. Listening to the radio in English for an hour a day is one such idea that was recommended to all students as a consequence of student feedback.

## VI. Conclusion

Preparing your own oral exam is not only a fairer way to assess the speaking ability of students than a written exam, but more importantly, it can lead to a change in student attitudes in the classroom. As a consequence of 'backwash', practicing English becomes real and students can see the point in talking to a classmate in English. Doing an oral test is also very beneficial for the teacher as it enables the actual productive ability of the students to be revealed in 'real time', rather than the students' ability to recognise English patterns (as in multiple-choice) or to compose English in the less pressurized format of a written exam, such as composition or dialogue writing. It also helps to focus the teacher's mind on achieving the goal of students speaking English.

Listening to students individually and providing each student with some feedback also makes the exam relevant to the students in a way in which merely receiving a grade cannot possibly manage. Although time consuming, it makes the process of testing constructive, and can provide motivation to the students to redouble their efforts in English.

If the teacher seeks the views of the students in asking them for anonymous feedback, then not only is the teacher likely to gain useful insights into how to improve the test in the future, but, it can provide information on students' perceptions of difficulties in speak-

ing English that might lead to a change in the teacher's classroom practice.

A well-prepared oral exam can have multiple benefits : it measures students' real speaking ability ; it can help improve students' attitudes to classroom work ; it can help the teacher focus on the aims of the course ; it makes the exam more relevant to the coursework ; and it provides invaluable information for the teacher on the difficulties that students are experiencing in English.

■Bibliography

Alderson, J. C.(2000) *Assessing Reading*, Cambridge : Cambridge University Press.

Bachman, L. & Palmer, A(1996) *Language Testing in Practice*, Oxford : Oxford University Press.

Bailey, K(1998) *Learning about Language Assessment*, Boston : Heinle and Heinle.

Brazil, D(1995) *A Grammar of Speech* Oxford, Oxford : Oxford University Press.

Buck, G(2001) *Assessing Listening*, Cambridge : Cambridge University Press.

Cohen, A.(1994) *Assessing Language in the Classroom*, Boston : Heinle & Heinle.

Cornbleet, S & Carter, R(2001) *The Language of Speech and Writing*, London : Routledge.

Douglas, D(2000) *Assessing Languages for Specific Purposes*, Cambridge : Cambridge University Press.

Heaton, J(1988) *Writing English Language Tests*, New York : Longman.

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

Kormos, J(1999) 'Stimulating conversations in oral-proficiency assessment : A conversation analysis of role plays and non-scripted interviews in language exams' in *Language Testing* 1999 16(2)pp 163-188.

Lazarton, A(2002) *Studies in Language Testing*, Cambridge : Cambridge University Press.

McNamara, T(2000) *Language Testing*, Oxford : Oxford University Press.

Shohamy, E(2001) *The Power of Tests*, Harlow : Longman.

Underhill, N(1987) *Testing Spoken Language*, Cambridge : Cambridge University Press.

---

### 英会話クラスのための口頭テスティングの重要性―口頭テストを行う際の枠組みの提案

Evans, David*　　　　Herbert, John
*国立看護大学校；〒 204-8575 東京都清瀬市梅園 1-2-1

【要旨】 本論文では，大学レベルの「英会話」の試験としてオーラルテストがなぜ最適であるかについて明らかにし，さらに教員がオーラルテストを行う場合に利用できる枠組みを提案する。　会話は，読み書きとは異なるスキルであり，オーラル以外の方法では会話力を測定することはできない。会話するには，聴力と会話力の双方が実時間で要求されるし，会話形式でないと的確には行われない。口頭言語の評価では主観性が避けられないので，口頭試験が決め手となるような試験に用いられることはないが，大学内では，教員が常時学生を評価できる環境にあるので，主観性も大きな問題とはならない。マルチプル・チョイスに代表されるような従来のテストと口頭テストの双方の利点および欠点について考察する。　オーラルテストの枠組みについては，教員が作成する場合に考慮したほうがよい問題点をはっきりさせる。例えば，習熟度に焦点をあわせるのか，それとも努力している実績をみるのかを考察したり，絶対評価にするのか相対評価にするのかを検討したり，現在行われているテスト形態をいくつか比較検討したりする。時間的な枠組みも提案する。教員は，学生の学習を促進し導くように個別にフィードバックを与えるとよい。また，テスト自体もよりよくなるように学生からのフィードバックを引き出すのが望ましい。

---