The "Why, What and How" of Oral Testing in an Oral Communication Program

The "Why, What and How" of Oral Testing in an Oral Communication Program

オーラルコミュニケーションプログラムにおける口述試験の必要性

Mark Hovane
David Svoboda

　教師の多くは、オーラルコミュニケーションのクラスにおいて、様々な形式を組み合わせることで評価を実施する。口述試験は、その構成や内容、測定方法が最も困難であると考えられているが、本稿では、大学のコミュニケーションコースにおいて生徒を評価するうえで、重要な構成要素であるとの視点から論じる。

　理論的枠組みや実践的考察を通して、本研究では、関西大学における English Communication 1a コースで、何故（why）、どのような（what）、そして、どのように（how）口述試験を実施すべきかについて述べる。また、本稿の目的は、Fluency-Driven Curriculum における教師への実践的支援を提示することである。

Introduction

Teachers use various forms and combinations of assessment for oral communication classes. Most teachers would agree that the most difficult form of assessment in terms of organization, content, and measurement is an oral test. Despite these difficulties, this paper advocates the use of oral testing as one of the major components of student assessment in university communication courses. Through an examination of both theoretical and practical considerations and research conducted in the English Communication 1a course at Kansai University, this paper will outline why, what, and how we can and should oral test. In addition, one of the purposes of this article is to provide practical assistance for teachers within a fluency-driven curriculum.

Research Aims

It is important for teachers who are proposing to include an oral test as a component of class assessment to have some understanding of students' previous test experience and attitudes towards an oral test. This information will impact on the degree to which the teacher should prepare the students for the test.

The research questions are:

1. Have you taken an English oral test?

2. How do you feel about doing an oral test?

3. How well do you feel you were prepared for the test by the teacher?

With these questions this research aims:

1. To ascertain an understanding of what percentage of students have previously

   undertaken an oral test for the purpose of informing the teacher in the relevant program to what degree they must introduce and explain the concept of an oral test for assessment purposes.

2. To identify student attitudes to an oral test which will influence the amount of pre-test preparation class time a teacher should give to students.

3. To assess how effective the various preparatory steps undertaken by the teacher were perceived to be by the students. The findings will influence subsequent actions by the teacher.

Literature Review

This review outlines prominent studies to support the present research and discussion in this paper with a particular focus on why teachers should oral test in a communicative program.

Testing oral skills has become more important as the role of speaking ability has become more central in language teaching. (Hartley & Sporing, 1999).

Oral assessment can be used to improve instruction and help students take control of their own learning. That is more likely to be accomplished when assessment is authentic and tied to the institutional goals of the program. (Bostwick & Gakuen,1995).

The institutional goals will vary from progam to program. In the Communication program at Kansai University the goals can be summarized thus:

- develop English skills necessary for effective communication in academic, business and personal situations
- give students a wide range of opportunities to practice and develop fluency in a variety of contexts

The most appropriate methodology for realizing these goals is communicative. This emphasis on communicative skills and fluency in the classroom calls for assessment to be oral based. We agree with Morrow's view (1979) that a language test should be proof that a learner can use the language and can demonstrate actual performance in real life situations. Written tests do not

reflect the goals of the Communication program . To use indirect testing techniques such as writing in a course which emphasizes the importance of spoken English in the classroom is inappropriate. Weir (1988)

One of the main benefits of oral testing is the 'washback' or 'backwash' effect. These terms describe the effect on teaching. Bachman (1990) highlighted that positive 'backwash effect' will result when the testing procedures reflect the skills and abilities that are taught in the course. Positive washback happens when students study and learn those things which teachers intend them to study and learn. (Hartley & Sporing, 1999).

This can have a positive effect on students. Thrasher (1984) calls this educational validity referring to the relationship among testing, study habits, test results, and course objectives in terms of the positive washback effects of the tests.

The very nature of oral testing and its immediacy can help teachers improve and modify their teaching methods and materials. As communication is the basic goal of the course, it is only by oral testing that accurate and informative feedback as to student achievement in-line with course goals can be measured and assist teachers in improving how they can achieve course goals.

Oral testing can increase motivation. (Antonio & O'Donnell, 2004)
It gives students the opportunity to show their communicative achievement in a communicative test. Conversely, Nitko (1989) found that using tests that inadequately linked and integrated with instruction can reduce motivation. Written tests in a communicative course would be an example of this.

An oral test effectively designed to favourably affect the student's perception of his/her speaking skills can increase both intrinsic and extrinsic motivation. Teachers can emphasize the communicative benefits of the test and the chance it offers to students to 'show off' the skills they have acquired over the course of the semester. During the course the extrinsic motivation of students is increased because they know at some point they will have to undertake an oral test and therefore improving their communication skills during class is essential. A self-perceived successful oral test can increase intrinsic motivation of a student and reduce the "I'm poor at English" syndrome that seems to pervade Japanese university student thinking

The oral test can not only be a reason for developing speaking skills, but also a means of achieving that goal. Through pair practice, classroom time allotted to test preparation and actual test performance, students are engaging in communication therefore developing their oral skills. Contrast this with preparation for a written test which will usually be a solitary activity with little or no oral production.

Achievement testing administered during or at the end of a course based directly on materials used in the classroom are the most appropriate and fair in an oral communication program. (Hughes, 1989).

Much research has been done on the construction and use of language tests (Hughes, 1989; Bachman, 1990; Brown 1996). In most discussions, writers focus on measurement strategies as either norm-referenced (NRT) or criterion-referenced (CRT). Brown (1995) is able to formulate a clear distinction between these two types based on the characteristics of the test itself and the logistics of testing.

To summarize briefly, NRTs measure general language proficiency while CRTs measure specific objectives. CRTs are more successfully used to motivate students by measuring to what extent they have achieved mastery of the learned/taught material. The Communicative Language Teaching (CLT) paradigm for language testing maintains that performance, rather than standardization should be the goal of measurement. Fulcher (2000) identifies four key words; validity, authenticity, performance and real life tasks as being important surrounding the CLT model for testing.

Subjects

All students are required to undertake English study in their first year and the students surveyed selected Communication 1a as their required English course.

Materials

There were three items in the questionnaire .Two items were Likert type 4 point scale questions and one item a multiple choice question and including an "other" choice which gave students the opportunity to write an answer not included in the alternative option. Percentage results were rounded to the nearest first decimal point.

Procedure

The questionnaire was given to 278 students in ten different classes on the completion of the oral test in July 2009 by their regular class teacher. The teacher explained the questions and answers in Japanese, prior to the students completing the questionnaire. Students were instructed that this was an anonymous questionnaire and to not write their names on it.

The questionnaire was collected by the teacher and read and interpreted by the researchers. The findings were then converted into percentages.

What should be tested

The two most common approaches to testing are proficiency and achievement. Proficiency testing is defined as being "independent of a particular syllabus, and provides a broad view of a person's language ability." (Beale 2008:2). If this approach is utilized for the Communication Course it is problematic for a number of reasons.

Firstly, the fact that classes are held only once a week over the two semesters, a total of thirty nine hours instruction time over the course of the academic year. Therefore, increases in levels of proficiency among students are going to be limited.

Secondly, students are not streamed into classes according to proficiency levels. Therefore, students who enter university with a lower level of proficiency are at a distinct disadvantage to some classmates who for a number of reasons beyond the scope of this study have higher levels of proficiency. Proficiency testing is therefore inherently unfair.

We propose a more valid and fair assessment test is achievement based and agree with Hughes (1989:13) "An achievement test only contains what is thought that the students have actually encountered, and this can be considered, in this respect at least, a fair test." The mastery of material taught or presented, progress made from some point to another. For example, new vocabulary learnt would indicate progress has been made. (Weir ,1988)

The fact that each Communication Course teacher produces and uses their own materials further supports the argument for achievement testing as a test should mirror what has been taught in the Communication Course context.

An examination of two oral test conversations recorded in a Communication 1b class in January 2009 between students illustrates the fundamental difference between the two testing approaches.

The target language that the teacher had taught in class included the following words and phrases;
Messy, tidy, morning person, night person, into.

Conversation 1

A: Are you messy or tidy?
B: Eto…… Maybe tidy?
A: Why?
B: ……….. Don't know.
A: I'm messy. My room always messy.

B: Are you morning or night person?

A: I like night.

B: When go bed?

A: 1. And you?

B: I'm a morning person. 7 wake up.

A: Are you into exercise?

B: E.. x..er..ci..se?

A: Ano….. run, swim?

B: No, lazy.

Conversation 2

A: Is your room clean or messy?

B: My room is always messy.

A: My room is always clean. I clean everyday.

B: Sugoi. My room is dirty. I… hate to clean.

A: Do you like exercise?

B: Yeah. I …do running.

A: Great! Do you get up early?

B: No. 8.

A: I'm morning person. Wake up at 6 because my house is far from Kandai.

An examination of the conversations would conclude that the speakers in Conversation 2 are more proficient speakers. There are less pauses, less use of Japanese, and better control of grammar. If the grading was based on proficiency they would clearly receive a higher grade than the students in Conversation 1. However, in an achievement based test the speakers in Conversation 1 did receive a higher grade because they used the target language more frequently and had mastered more vocabulary items from classroom instruction.

Students need to be made aware of the achievement based testing approach of the teacher from the beginning of the course and this can be emphasized using the above dialogues to illustrate this.

PREPARATION

Students need to be properly prepared to undertake an English speaking oral test. Student

reactions and questions indicate that the prospect is intimidating to the majority of students so they must fully understand the process and what is required of them. A survey of students was conducted in ten Communication 1a classes in Semester 1, 2009.

Question. 1 Have you taken an English oral test？

Table 1 English oral test experience.

	$N = 278$	%
a) Junior/Senior High School	33	11.8
b) Eiken	74	26.6
c) Other	6	2.2
d) No	165	59.4

Question. 2 How do you feel about doing an oral test？

Table 2 Attitude to the English oral test.

	$N = 278$	%
a) Not nervous	12	4.3
b) A little nervous	20	7.2
c) Nervous	62	22.3
d) Very nervous	184	66.2

The results indicate that for the vast majority of students, a Communication Course oral test will be the first time they have undertaken an English speaking test and that they approach it with considerable trepidation. Teachers play a crucial role in student performance in the test, as their performance depends to a general extent, on how well they understand what is expected of them and the conditions under which the test is taken. The teacher can undertake a number of practices that will make the prospect of the test far less intimidating and enhance student performance.

Video

To properly prepare 2009 Communication 1a students for their spring semester test the writers showed a video of a pair of students undertaking a test in autumn semester 2008. From this the students were able to observe a number of facets of the test. Firstly, the physical layout of the testing environment. Secondly, the type of interaction they are expected to undertake. Finally, the role of the teacher in the test. This visual element to the prospective test has been very well received by students and the ten minute video answered many student questions.

Revision class

A number of program teachers allocate one lesson to revision. Students are free to revise and prepare for the test. Classroom observation has found students use this time effectively to prepare for the test.

The revision class serves as a means of 'tying up' all that has been taught in class over the semester. During the class the student English output is extremely high and we consider a revision class to be an effective component of the course for all the abovementioned factors.

Pairs / Groups

Most teachers in the current program favour testing students in pairs or small groups. Research has found that the relationship between test takers in a group can affect performance. (Scott, 1986)

As far as possible, it is preferable to allow students to choose their own partners or group members who they have become familiar with and more comfortable in speaking English over the course of the semester.

Teacher Actions

Students naturally are at their most nervous at the beginning of the test. The first minute should be given to teacher – student interaction in a relaxed manner to put students at ease for the more formal component of the test. A number of behaviors by the teacher will only increase student tension. Teachers should avoid looking disinterested in what students are saying and indicate interest by nodding, smiling and generally looking alert. Mistakes should not be corrected and notes should not be made during the test as such acts can only increase student nerves.

Question 3. How well do you feel you were prepared for the test by the teacher?

Table 3　Preparation for the test

	N = 278	%
a) Not well prepared	14	5.1
b) Prepared	22	7.9
c) Well prepared	161	57.9
d) Very well prepared	81	29.1

The survey results indicate that the great majority of students considered themselves to have been well or very well prepared by the teacher by means of a video example of an oral

test, a review lesson and actions by the teacher at testing. Clearly, utilizing the abovementioned preparation tools will allow students to perform to the best of their abilities.

How we can oral test

The Unique Role of Oral Testing

An oral test cannot be treated just like any other more conventional test. Usually a test is seen as an object with an independent identity and purpose, with the people taking the test being "reduced to subjects whose only role is to react to the test instrument." (Madsen, 1983:159). Having deemed achievement based oral testing to be the most appropriate method of assessment for an oral communication course; how can we as teachers construct a tool of maximum utility for both the test taker (student) and the test developer (teacher) alike? It is an interesting irony that although teaching speaking skills is clearly at the forefront of many university communication courses, the assessment of those speaking skills often lags behind . Most researchers in the field agree that "the testing of speaking is the most challenging of all language tests in terms of: preparation, administration and scoring." (Madsen, 1983:147).

The reverse is true in an oral test where the people themselves are more important than the test and the interaction between the participants is fore-grounded. Thus communicative testing has a problem which is not shared by regular psychometric testing which sees language more as a series of discrete structures rather than as a means of communication. Fundamental to the set up of the most useful test vehicle, is developing an awareness of "WHO" it is for, as this will guide any attempts to structure the test. It is clear for many reasons that it is the students more than the teacher who benefit most from the administration of an oral test. As has already been mentioned, we believe that the primary rationale for oral testing is to be motivational for the students in the context of a first year undergraduate program at a Japanese university.

Carroll (1981:8) highlights the unique role of oral testing and stresses that language be taught and tested according to the specific needs of the learner. In the context of Kansai University's first year Oral Communication program, if we define those needs under the heading of communicative competence, then we must be prepared to accept the trade-off between reliability and validity. (Underhill,1982). However this decrease in reliability might be compensated for by a positive increase in authenticity because the test would reflect the curriculum content. For a typical, "achievement based, end of semester oral test, in measuring the improvement in communicative competence, the difficulty of administering a "pre-test" and lacking other reli-

able indicators of initial language ability make it obvious that it is essentially hard to quantify results. However, quantification of results is not the main purpose of the exercise. The general trend of the results is that there is a very low self-perception of oral abilities among students who generally enter university classes with very undeveloped practical skills and on the other hand, major confidence barriers and anxieties about speaking English. A prime aim of the Communication program at Kansai University is to address these problems and to help the students perceive themselves as successful producers of what amounts to a vast reservoir of largely untapped, passive knowledge. Anxiety can be a strong inhibitor of performance, particularly in oral tests and every attempt should be made to help reduce any factors that contribute to it, by prioritising student test preparation through various means.

Transition from Norm-referenced to Criterion-referenced Testing

It can be useful to recognize that Japanese students emerging out of an examination based secondary education, tend to be particularly "test-driven", therefore it can be useful and effective to use the instrument of a "test" itself as a means of achieving course goals. Griffee (1995) speaks of student expectations of a final examination as a means of evaluating the seriousness of a course. The big difference between secondary and university assessment is that at university, the test result is not 100% of the available score.

According to the guidelines for grading given to teachers in the Kansai University Oral Communication program, a maximum of 30% of marks will be awarded by assessment and evaluation tools. Therefore, continuous, formative assessment is the basis of the student grade. Interestingly, in terms of transitioning from a learning culture of almost completely summative assessment to one that includes a large component of continuous and formative assessment, while the reality should be overwhelmingly clear as to favour the "week by week" progress of the student, the residually fossilized perception is that "the test" is the most important thing. Given this history, another justification for an oral test is to marry with students' expectations because if there was no "test" structure at all and only continuous assessment was used, it is highly likely that the course would not be seen as valid by the students who have been taught that they need to "see results". It is desirable therefore, that the students also create a paradigm shift in the way they think about assessment. Coming from a norm-referenced, exam driven assessment style that has characterised their secondary education, to a criterion referenced style is a major change.

Oral Test as Motivational Tool

Given the entrenched nature of the "test mentality", the teacher can almost "subversively" use studying "for the test" as an effective motivational tool. The requirement of having to "perform" in English in the test, means that students and teachers must concentrate on practicing "performance" in lesson time. This leads to a second effect, the realization that skills development is a gradual rather than an "all or nothing" process, with students beginning to see that if progress is to be measured in terms of performance, then leaving study to the last moment is not an effective strategy. If the teacher is able to simultaneously create and use an authentically communicative oral test as a motivational tool, while impressing upon the students the explicit nature of the university based grading system, there is every possibility that the students' paradigm of learning culture will start to shift. The assessment itself should be as student centered as possible so that in addition to being a grading tool, it provides students with a structure that allows them to be more involved in their own learning. As the students come to transform their thinking about assessment, they come to re-invent its place and see it as an essential part of the learning process and not just something to be added on at the end of a series of lessons.

Although we can conclude that administering an oral test is primarily for the students' benefit, it is also useful for the teacher in monitoring how students have achieved course goals and in providing useful feedback for curriculum review even if questions of reliability remain unanswered.

Challenges of Oral Testing

Notwithstanding the obvious benefits to both students and teachers alike, we need to take into account the multiple challenges that face the topic of oral testing. It is generally perceived that oral testing is a difficult and perplexing problem for language teachers.（Nagata, 1995）. Problems include: practical concerns of administration: designing productive and relevant speaking tasks, deciding which criteria to use in making an assessment and how the selection and weighting of these criteria depend on the exact circumstances by which the test takes place, not to mention the problem of consistency with different testees on different occasions. As Bachman（1990）has pointed out, test methods also have an important effect on test performance. Facets of test methods that might affect performance include the testing environment as well as test rubric etc. When test performance starts to be affected by factors other than the abilities being measured, this might lead to a compromised validity of score interpretations. How can a teacher accommodate all these difficulties and still come up with a valid test of oral

production?

Guidelines for Oral Testing

In describing the background of how to set up an effective speaking test, Beale (2008), posits a framework or set of guidelines to make the assessment less arbitrary. Guidelines include: practicality, validity and reliability. Practicality is concerned with the logistics and ease of administration of the test given the constraints of time and the number of students to be tested. Essentially, validity concerns the question of: "how much of an individual's test performance is due to the language abilities we want to measure?" (Bachman, 1990:161). Reliability deals with the extent to which the results are quantifiable and objective and the degree to which we can therefore depend on the test results to be consistent. Weir (1990) also identifies an inevitable tension between validity and reliability, arguing that it is sometimes essential to compromise a degree of reliability to enhance validity. In moving from norm-referenced multiple choice tests to freer productive tests, it is generally accepted that reliability will be inevitably reduced. However, when it comes to speaking tests, it's necessary to make a distinction between score reliability and task reliability. Task reliability is directly proportional to the degree to which students believe that the test measures their speaking ability by employing valid speaking activities. Amongst these factors, Nakamura (1995) describes validity as the "single most critical element" in constructing tests. Specifically, we need to determine which types of validity are the most important for an oral test which attempts to measure achievement rather than proficiency.

Face, Content and Educational Validity

According to Davies (1990), an achievement test should have both face and content validities in particular. Nakamura (1995) also makes a claim for educational validity.

Face Validity

Firstly, many designers of communicative tests regard face validity as the most important of all types of validity. Face validity is concerned with the appearance of the test to the teachers and learners who use it and the degree to which it is considered fair. A test with a high face validity will maintain students' motivation. Direct speaking tests like pairwork-interview tests have much more face validity than indirect tests of speaking skills such as paper-and-pencil tests. Underhill (1987) suggests that questioning students after a test has been administered, is a good indicator of how "reasonable" it is. High face validity would also explain

why students tend to be excited about taking this type of "authentic" test, notwithstanding the fact that it is largely unfamiliar.

Content Validity

For an achievement based oral test, content validity is also a vital concern. This measures to what extent the test items mirror the language skills and structures contained in the syllabus itself. It is generally understood that tasks are less important than the match between classroom and test grammar and vocabulary. Teachers must be careful to use test tasks that incorporate oral course objectives.

For Kansai University's Oral Communication program, it is useful to think of the syllabus as containing a continuum of objectives that range from broad to narrow.

An example of a very general goal might be written as:
- to give students a wide range of opportunities to practice and develop fluency in a variety of academic, business and personal situations.

An example of some more specific goals from the English Communication 1a syllabus guidelines are:
- students will be able to learn:
- the basic structure of a conversation, i.e. how to start, continue and finish a conversation.
- how to keep a conversation going by asking follow-up questions.

From the same set of syllabus guidelines in the category of language areas:
- students will be able to learn:
- topical vocabulary (ex. friends, personality, food, etc.)
- how to form questions (i.e. basic patterns)

Underhill (1987: 106) maintains that "content validity can be assessed by comparing the kind of language generated in the test against the syllabus.". Therefore to maintain content validity, the design of the oral test must address and be driven by these syllabus goals directly.

Educational Validity

In addition to face and content validities, Nakamura (1995) cites educational validity as being crucial to effective oral testing. This opinion follows Thrasher's (1984) view that content validity was not sufficient from the standpoint of the appropriateness of teaching. Educational validity involves the interdependence of: testing, teaching, study habits and test results from the point of view of a positive washback effect on the motivation of students as previously discussed in this paper.

In the Japanese university context, educational validity mirrors the change in student study habits from the secondary to the tertiary school setting. That is to say; from focusing on grammar based study to a more communicative (listening and speaking) approach. A high level of educational validity would suggest that students start to focus on the productive aspects of their language skills and pay more attention to context. Teachers would also ensure that their syllabus maximized opportunities for communication in "real life" situations.

In summary, in considering that one of the main purposes of the "end-of-semester" oral test is deemed to be motivational, we can see that the non-empirical forms of face, content and educational validities are all vital aspects to be incorporated in the design of an effective oral test.

Best Format for Oral Testing

Having taken into account the important issue of validity, how does the teacher choose the best test format? Weir (1988:82) states that "communicative testing is purposive, interesting, motivating, interactive, unpredictable and realistic."

One of the key characteristics in assessing interactive language is that by definition there is another person taking part. Underhill (1987) states that the person to person aspect is vital. Thus both productive and receptive skills are being tested. Kitao (1996) mentions that in assessing productive skills, the focus tends to be on appropriateness rather than grammatical accuracy, while conversely for receptive skills, the focus tends to be on understanding the communicative intent of the speaker.

Semi-structured Conversation Test

Of the many kinds of oral assessment task that can be used in an end-of- semester oral test, the writers of this paper suggest a "conversation style" test between two learners in which the teacher acts primarily as "listener" during the test and "assessor" after the test has finished. This type of test frees up the cognitive resources of the teacher to be able to pay closer attention to the production of each student as well as allowing students a longer time to interact. This style of oral test is not dissimilar to the interaction task that Weir (1990:78) terms an "information gap student-student".

This "conversation test" is semi-structured in that students are expected to utilize the basic components of a conversation (beginning/middle/ending) structures while allowing some freedom or level of unpredictability especially if the teacher withholds the actual topic of conversation until just prior to the conversation test itself. Naturally the range of topics

selected as a basis for the conversation test will be taken from the syllabus and will be charac-terized as "high-interest" to first year undergraduate Japanese university students to ensure student motivation for communication. One of the advantages of this method is the increased validity as a test of "real life" oral skills, but at the cost of reliability of measurement due to some of the unpredictable nature of the testees' responses. (Underhill, 1987)

This semi-structured style builds on a basic architecture of patterned responses that incor-porate learned chunks of language as a framework, while still allowing an unpredictable element of interaction management and negotiation of meaning to influence the exchange. The combination of routine and improvisational elements within the structure of the test, helps keep the interaction more authentic and goes some way to avoiding the stilted effect of rote memorized textbook dialogues. "Weaning" the students off the overly familiar technique of rote memorization, also helps to prevent the occurrence of the 'trance effect" in which students deliver a completely memorized quasi-monologue of desperately learned gambits with little or scant attention paid to their partner in the simulated "real life" setting. The secondary pitfall of the "trance effect" is that should the student not remember perfectly what they had prepared to say, the tendency is to panic and become "frozen in the headlights". Typically this "paralysis" or breakdown in communication would likely result in a negative washback, reinforcing a sense of failure. Therefore the semi-structured conversation test allows learners to make the transi-tion from a familiar memorized dialogue to a more improvised interaction.

It should be mentioned that this is just one of many possibilities of a criterion-referenced assessment and ultimately it is up to individual teachers to administer a test that most closely reflects their own curriculum. Whatever the form decided, a defining characteristic should be that actual performance of relevant tasks be required of test takers (students), rather than more abstract demonstration of knowledge such as that required by tests of ability.

Assessment Criteria

Having decided the test format, the next consideration is how to determine assessment criteria and an appropriate scoring rubric.

There are many ways to specify performance criteria for the criterion referenced oral test. Brown (1996) posits that the assessment criteria need to be related to the actual purpose of the test. In a criterion-referenced, achievement based test, course objectives, topical vocabulary and structures taught during the course would constitute the basis of the assessment criteria. After specifying criteria, it is still necessary to determine which of the categories are more important and "weight" them accordingly.

Scoring Rubric

Having specified assessment criteria, the next task for the assessor is to develop appropriate scoring procedures. (Madsen, 1980). In an attempt to make the marking of subjective oral tests more consistent, Bachman (1990) suggests a system of objectified scoring. Involving the development of a classroom specific rating scale as a rubric for grading student performance on oral tests, this process helps to make marking explicit and is therefore a more transparent alternative to completely impressionistic marking. Even so, the problem exists that ratings always involve subjective judgements in the scoring process. The main problem is making explicit assumptions regarding oral communicative competence and applying these theoretical constructs in assessing the actual samples of student elicited performance for each oral test. Underhill (1987) suggests parameters that might be appropriate for a criterion-referenced achievement test might include: (1) fluency of speech, (2) vocabulary appropriateness and complexity, and (3) flexibility. Each parameter might be evaluated according to the application of a Likert-type scale where (1= poor) through to (5) being excellent. These parameters could be continually adapted through repeated usage. It should be concluded that each assessor will determine his/her unique descriptive assessment criteria. As a rule of thumb, Underhill (1987) claims that fewer levels make assessment easier and reliability higher.

Further Considerations for Scoring

An achievement test that is criterion referenced will assess students individually on their achievement of learning outcomes. Score distribution has a direct correspondence with learning success and therefore it is possible in theory for all testees to receive 100%. This distribution would not present a problem in a system where only up to 30% of the available grade is determined by the results. This system of marking is in contrast to a norm-referenced test which would aim to rank students on the basis of making distinctions between their performances. Once the assessment criteria and scoring rubric have been decided, we advocate that students be informed in advance of the test so that they might take greater responsibility for their achievement. These scales also offer the chance to consciously incorporate course objectives into their tests, thus maximizing the possibility of a positive washback effect.

To be of greatest benefit, we also recommend giving performance feedback immediately after the test is taken, either orally or in the form of a brief written comment.

While admitting that oral testing is an inexact science, what we have outlined, are important steps in demystifying what is inherently a subjective evaluation process. We should also remember that oral testing is highly influenced by internal and external factors that are inde-

pendent of language use. Thus we can see the challenge of creating a good oral test involves minimizing external factors and creating an environment in which the testee can give of their best ability.

CONCLUSION

One of the points this paper has tried to illustrate is that the use of criterion-referenced, oral tests, focusing on communicative competence for the Kansai University's Oral Communication program will have a beneficial washback effect of ensuring that the courses focus on the means of promoting oral skills. If the stated goal of the program is to develop spoken English then the incorporation of an oral test into the present testing system is to be highly recommended.

By administering tests which not only assess the level of oral skills but also assist in the very improvement of these skills, the issue of test driven learning is given a positive aspect since the way to pass the test is to participate in the classes and to give the oral skills the time to grow. By doing this, the student is acquiring beneficial learning habits and the test is therefore fulfilling more than one pedagogical aim.

In attempting to answer the "why/what/who and how" of oral testing in the classroom, we can conclude that we need a test which mirrors what has been taught (high content validity), that is learner centered incorporating high face validity, and that has high educational validity as well. It should be seen as working for students rather than against them. Given the limited communicative context in which English has been experienced in the Japanese secondary school setting, the real success of this test would be judged primarily by its effectiveness in favorably influencing the student's perception of his/her spoken abilities, since a self perceived improvement would result in increased confidence when using the language and would positively affect motivation to continue learning.

REFERENCES

Antonio, J. & O'Donnell, K. (2004) Using Criterion Referenced Assessment Toward a Reorientation Student Motivation, *The Language Teacher*, 28 (3) 19–23

Bachman, L.F. (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press

Bachman, L.F., & Palmer, A. (1996) *Language testing in practice*. Oxford: Oxford University Press

Beale, J. (2008) Assessing interactive oral skills in EFL contexts. Retrieved March 10, from http://www. jasonbeale.com/essaypages/assessment.html

Bostwick, R.M. & Gakuen, K. (1995). Evaluating Young EFL Learners: Problems and Solutions. In Brown, J.D. and Yamashita, S.o. (eds), JALT Allied Materials Language Testing in Japan. Tokyo: The Japan Association for Language Teaching: 57–65

Bray, E. (1998) First Year English Students Backgrounds, Interests and Motivation: Before Instruction. *Yokkaichi University Journal of Environmental and Information Sciences* Vol 1 (1, 2)

Brown, J.D. (1995). Differences between norm-referenced and criterion-referenced tests. Cited in Brown,J.D. & Okada Yamashita, S. eds (1995) *JALT Applied Materials: Language Testing in Japan*. Tokyo: The Japan Association for Language Teaching. 12–19

Brown, J.D. (1996) *Testing in language programs*. Upper Saddle River, N.J., Prentice Hall Regents

Burden, P. (2002). Retrieved November 8 from http:// www.jalt-publications. org/tlt/articles

Carroll, B. (1981) *Testing communicative performance.* Oxford: Pergamon

Davies, A. (1990) *Principles of language testing*. Oxford: Blackwell

Fulcher, G. (2000) *"The "communicative" legacy in language testing." System 28(4)* 1–15

Griffee, D.T. (1995) Criterion-referenced test construction and evaluation. Cited in Brown, J.D. & Okada Yamashita, S. eds (1995) *JALT Applied Materials: Language Testing in Japan*. TJapan Association for Language Teaching. 20–28

Hartley, L. & Sporing, M. (1999). Teaching communicatively : assessing communicatively? Language Learning Journal, 73–79.

Hughes, A.C. (1989) *Testing for language teachers*. Cambridge: Cambridge University Press.

Hughes, R. (2005) The Need for Oral Proficiency Testing As A Motivational Tool in Japanese Universities. *Journal of Regional Development Studies*

Kitao, K., & Kitao,S. (1996). Testing communicative competence. *The internet tesl journal*, *2*(5). Retrieved June 9, 2009 from http://iteslj.org/Articles/Kitao-Testing.html

McVeigh, B.J. (2001) Higher education, apathy, and post-meritocracy. *The Language Teacher,* Vol 25 No 10, 29–32

Madsen, H.S. (1983) *Techniques in testing*. New York: Oxford University Press

Morrow (1979) Cited in Weir, C.J. (1990) *Communicative language testing*. London: Prentice Hall

Nagata, H. (2005) Testing Oral Ability: ILR & ACTFL Oral Proficiency Interviews. Cited in Brown, J.D. & Okada Yamashita, S. eds (1995) *JALT Applied Materials: Language Testing in Japan.* Tokyo: The Japan Association for Language Teaching.

Nakamura (1995) Cited in Brown, J.D. & Okada Yamashita, S. eds (1995) *JALT Applied Materials: Language Testing in Japan*. Tokyo: The Japan Association for Language Teaching.

Nitko, A. (1989) Designing tests that are integrated with Instruction. *Educational Measuremen*t. Ed. London: Longman.

Redfield, M. & Larson, S. (1995) How University Faculties Differ: A Look at Communications Mass survey Data. *Annals of the Research Center for General Education*, Kansai University, 22(1) 41–63

Scott, M.L. (1986) Student affective reactions to oral language tests. *Language Testing* 3, 99–118

Thrasher, R.H. (1984) Educational validity. *Annual Reports*, International Christian University, 9, 67–84

Underhill, N. (1982) Cited in Heaton, J.B. (Ed.) (1982) *Language testing*. Hayes, Middx.: Modern English Publications

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

Weir, C.J. (1988) *Communicative language testing*. London: Prentice Hall