The State of Speaking Assessment in Freshman English Classes
Adrian Wagner, Asia University

Abstract

Freshman English Classes are compulsory for all first year students at Asia University. While the classes have no fixed format, the general purpose is to increase communicative ability of students. Currently, there is no policy or standardized format regarding assessment for Freshman English classes. This study investigates how instructors, on an individual basis, are designing and implementing assessment of speaking skills in these classes. It assesses which methods of assessment are being used, instructors’ reasons for choosing these methods and their attitudes towards assessment of speaking in general. The results reveal that assessment of speaking ability is something that instructors in the Freshman English program are approaching in many different ways. It also shows that despite there being no set criteria, individual instructors attempt to ensure that through various methods of assessment, speaking ability and achievement are adequately represented in allocation of final grades.
The State of Speaking Assessment in Freshman English Classes

The Freshman English program is a compulsory course for students in all departments at Asia University, spanning over two semesters. For most departments, it entails five, 45 minute classes per week, all of which take place in either the first or second period. For students it is often the first experience of communicative English language classes. For most instructors in the Center for English Language Education, Freshman English compromises if not the majority, then at least a significant portion of their teaching and preparation time. So what is the outcome of this expenditure of so much time and effort? Students will receive a grade of D (failure) or a grade of C, B, A or S and their compulsory credit points. What this grade represents is in general, and ideally, a combination of attendance, effort and achievement in this class.

Attendance is easily quantifiable. Effort is not. And to what extent is achievement in a language class quantifiable and how can we do that? In communicative language classes, where the objective is increasing students’ speaking and listening abilities, how can we create assessment that is fair, comprehensive and corresponds with our class content and classroom realities?

This paper attempts to investigate how instructors of the Asia University Freshman English program are dealing with these problems. Specifically, it explores which methods of assessing speaking instructors are using and questions whether speaking ability and achievement are under-represented in their methods of assessment.

The Place of Speaking Assessment in Communicative Freshman English Classes

At the end of the Freshmen English course, students must be given a grade. If this grade is to mean something and not be completely arbitrary, some form of assessment is arguably required. In Freshman English classes, 40% of the grade is allocated to attendance and participation. How this is interpreted and allocated is up to each instructor. Furthermore, instructors are responsible for deciding what the other 60% of the grade entails and
communicating what is expected of students to obtain that 60%. Under the current circumstances, students are essentially being assessed and given a number and corresponding letter as proof of their achievement over each semester.

The expectation of giving a final grade is not the only factor necessitating assessment. It is difficult to imagine formalized education without some form of assessment. According to Allen et al. (2011), “Assessment is an integral component of learning and teaching. It refers to all processes employed by academic staff to make judgments about the achievement of students in units of study and over a course of study” (p.1).

Although often overlooked, one’s forms of assessment make a big statement to students. Consider the common question asked by students of all disciplines all over the world, “Will this be on the exam?” In the context of the class, assessment methods emphasize what is important and should motivate students. They should be fair, demonstrate that standards are being maintained and essentially determine who will pass and who will fail. Lastly, assessment should ideally match the content of the course and the learning outcomes that we as educators have set for our students. As O’Farrel notes:

…while it might be a wholly appropriate assessment method when testing for knowledge and evaluation of, say, “the theme of Irish tradition in contemporary reviews of 1930’s art exhibitions”, it would be less appropriate to set an essay to assess whether a student had mastered a particular brushstroke. (2005, p.5)

Speaking is a practical skill like painting, as such, it must be examined through practical methods of assessment. Also, not unlike painting, passing judgment upon speaking is full of inherent subjectivity and complexity. Although debatable, language is generally not regarded as being as easily testable as mathematics and sciences or even other humanities such as history.
Speaking is arguably the most difficult of the four skills to assess. Writing about theoretical complexities in preparing speaking tests, O’Sullivan (2008) states, that test writers have a lot to think about, “most notably construct definition, predictability of task response (task description), interlocutor effect, the effect of characteristics of the test-taker on performance, rating-scale validity and reliability, and rater reliability” (p.1).

And of course, there are other things to consider such as time and logistics. Consider an instructor implementing a one-on-one speaking test in the Freshman English classes at Asia University. What are the other 19 students doing? How much time is needed for each student? How many class days will this take? Can using this much class time be justified?

Forgetting the theoretical and practical complexities of assessing speaking, testing is subject to constant debate in Japan. There has been considerable criticism of the teach-to-test tradition and the constraints that places on teachers and students in junior high schools and high school. It comes as no surprise that when such students reach university, they are more than familiar with taking English exams. “The infamous juken jigoku “examination hell” places students under tremendous pressure and the competitive nature of entrance examinations cause schools to focus on English material that is likely to be seen on the exams” (Hiramoto, 2013, p. 231). While a listening section was added to the National Center Examination for University Admissions used for entrance to public universities in 2006 (Guest, 2008), English exams generally focus on advanced grammar knowledge and reading comprehension.

Many commercially available English tests such as TOEIC and TOEFL contain a speaking section (Wagner, 2015) but most students have not taken any of these prior to their matriculation at Asia University. Despite the prolificacy of English exams in the Japanese education system as a whole and the MEXT stated goal that, “… a junior high school graduate is to be able to converse in plain English, including greetings and corresponding dialogues” (Hiramoto, 2013, p.232), there is no standardized way of assessing this. Also, while some
research (Akiyama, 2003) suggests that teachers at junior high schools and high schools in Japan are administering speaking assessment on an individual basis, there is a high possibility that students in Freshman English classes will be having their speaking ability directly assessed for the first time.

In this regard, instructors of Freshman English at Asia University are in a position of apparent freedom given that there are no external, standardized tests or students’ experiences and expectations influencing their assessment design.

This study is based on the premise that we design tests based on four factors:

1. The course content and textbook
2. Our own philosophy and experience
3. Our learners
4. Practical considerations such as time and student numbers.

These factors are based on reading about speaking tests and informal discussions with educators in various contexts. They also take into account that while in CELE, we may operate with some factors, in common, for example class lengths and textbooks, we all have different backgrounds and no two classes or learners are exactly the same. The following study is designed to investigate how instructors at CELE juggle these factors and produce speaking assessment tasks.

The Method of Study and the Respondents

There were 24 respondents to the survey. All are full time employees of the university who have at least a Master’s Degree and have demonstrable teaching experience. At present, all of the respondents are teaching at least one section of Freshman English from the departments of Business Hospitality, Business Administration, Economics, International Relations or Law. The majority of Freshman English classes last forty-five minutes and are held either four or five times a week.
The survey was made using Google forms and offered the condition of anonymity. It had a range of multiple-choice and open-ended questions to enable data subjectable to both quantitative and qualitative analysis. The survey was intentionally forwarded to instructors in the final week of the university semester when final assessments were being completed and final grades were being compiled. Therefore, assessment was fresh in the instructors’ minds while completing the survey.

As the data set was relatively small (only 24 respondents) and most of the questions were open, requiring a written response rather than a selection, the data has not been subjected to much statistical analysis. Where relevant, responses from instructors have been included in the survey to enhance the qualitative nature of the data presented.

**Analysis and Discussion of Results**

**General Goals and Practices**

The first two questions focused on how much importance instructors’ placed on speaking skills, and the weight of speaking activities in their respective classes.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus on Speaking in Freshman English Classes</strong></td>
</tr>
<tr>
<td>1. Would you say that the development of speaking skills is a central focus of your freshman English class?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2. About what percentage of class time is used for speaking activities?</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
According to the table it is evident that the vast majority of instructors agree that speaking is a central focus of their Freshman English classes. Furthermore, the majority of instructors estimate that at least half of their class time is devoted to speaking activities. Unsurprisingly, instructors who claimed speaking was not a central focus of the Freshman English class were more likely to spend less class time on speaking activities.

To make a point of comparison for later questions about speaking tests, questions 3 and 4 focused on what kind of other assessment activities are given to the students.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Nonspeaking Assessment in Freshman English Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you give a test that includes listening?</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you give a test that includes grammar, vocabulary, reading and writing?</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

Given these results, it is safe to assume that the vast majority of instructors assess non-speaking skills in their Freshman English classes.

At the beginning of each semester, instructors write a syllabus for each class which includes information as to how the class will be graded. Question 5 required respondents to state whether they assessed speaking and the relative weight of speaking outlined in their syllabi. Respondents were asked to write freely for this section. Their responses are presented in Table 3. The scores do not add up to 24 as some teachers answered differently for different classes, depending on department and how many class meetings per week. All responses have been included in the table, therefore, numbers in the results in Table 3 represent separate syllabi rather than instructors as in Table 2 above.
Table 3

<table>
<thead>
<tr>
<th>No</th>
<th>Yes (as part of participation)</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>45%</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Question: In your allocation of grades (on your syllabus) is there a percentage allocated for speaking?

The results of this particular question were far less consistent than the previous ones, representing the different approaches taken by instructors in designing their courses and allocation of grades.

**Methods of Assessing Speaking**

Question 6 investigated which methods of assessing speaking were employed by teachers. Again, the number of responses does not reflect the number of respondents to the survey as some instructors reported using more than one type of speaking assessment in their freshman English classes.

Table 4

<table>
<thead>
<tr>
<th>Only as part of participation</th>
<th>Systematic Ongoing Assessment</th>
<th>Presentation</th>
<th>Speaking Tests</th>
<th>Narrated Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>6</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>

Perhaps due to poor survey design, there are some inconsistencies between the results for instructors who suggested they used ongoing assessment and those who reported to have only included speaking as part of the 40% attendance and participation points.

The first caveat to these results is that the first option, “I only include speaking in the 40% allocated for attendance and participation,” could be interpreted as an unformulated method of ongoing assessment. Note how the word “only” is included in the response. This survey allowed respondents to differentiate between the two, so the results are displayed accordingly.
Secondly, of the instructors who suggested they were using ongoing assessment, (according to their responses to Question 5) the majority of them (excepting two) did not seem to include a specific grade percentage for speaking as assessed through ongoing assessment on the syllabus, so it is assumed that the ongoing assessment component of their speaking grade was included in the 40% attendance and participation points.

Ongoing Assessment

A number of instructors reported to be using ongoing assessment to assess speaking, or seemed to be doing so by default by including speaking as part of the score for attendance and participation. Ongoing assessment (also referred to as continuous assessment and performance-based assessment) is seen as an alternative to standard exam style testing. It is usually formative (in that it occurs many times, over the course of study) rather than summative in nature (taking place at the end of the course).

While summative assessment has its advantages in that it is easier to administer to large numbers of students and is more standardized, there is significant support for formative assessment. “Continuous assessment is ongoing and is based on observations of what students are doing. Examinations are one way of assessing learners but they are only a ‘snapshot of the learner.’ Continuous assessment is like having many ‘snapshots of the learner’ (American Institute for Research 2003, p.7).

The purpose of ongoing assessment is to collect information about the students’ performance over time, rather than at a single time. This can be seen as fairer or more consistent.

Summative assessment is teacher-centered and its essential goal is to determine grades. In an oral class where grades are determined by oral tests, the results are essentially subjective as the student’s performance is determined by single observations whether using an assessment scale or linguistic descriptors (Gobert, 2003 p.64).
Another advantage of ongoing assessment is that while the instructor is assessing the students’ learning outcomes on a regular basis, they can use this information to constantly reform and refine class content and teaching practices. This is also applicable to students as they have an opportunity to receive feedback and guidance on an ongoing basis.

The following are instructors’ descriptions of their style of ongoing assessment. The first refers to assessment data collected during the course of regular classroom activities:

*I collect several random samples and add them all together. At various times during a class (but only once per class) I take a look at who is speaking English. I subjectively evaluate participation in class and award one of four levels 1/4, 2/4, 3/4, or 4/4 points. These all get added up at the end (I try to take between 15 and 25 measures).*

A different instructor refers to ongoing assessment in conjunction with formative assessment of presentations:

*I emphasize key aspects of effective presentations. I give students two practice presentations where further, critical feedback is offered. I then provide rubrics with a clear explanation of how these points relate to effective speaking, and how I will evaluate their performance.*

The assessment is formative in nature as it is initially more concerned with providing feedback and guidance than a grade and occurs several times.

When justifying why they chose this method, instructors referred to the fairness of ongoing assessment, it being both a natural form of assessment and motivating for students. For instance, here are responses from three different instructors:

1. *Language learning is a fluid and evolving process - I want to try to capture as many of those moments as I can in order to support my evaluation of the student’s speaking skills.*

2. *In a class of 20 plus students, it seemed complicated to assess speaking on a more individual basis. I was also worried about unintentional bias against shy, quiet students.*
3. It seemed the most effective way for encouraging students to participate in class.

4. Because there are many factors that can make a student very nervous if they just focus on a speaking test. This way they do not know or feel they are getting evaluated.

Although provided by four different instructors, these statements aptly encapsulate the positive aspects of ongoing assessment.

**Presentations**

Teachers using presentations had significantly varied tasks and used different grading criteria, so much so that putting it into a table would be difficult. Lengths ranged from one to ten minutes, and teachers reported assigning the task to individuals, pairs and small groups. Grading techniques were also quite variable, according to the type of tasks and themes of the presentations.

While the survey did not ask directly if the methods of grading the presentations were analytic or holistic, the vast majority of instructors assigning presentation tasks use what would be closer to analytical than to holistic grading criteria. A selection of some varied and interesting assessment criteria are provided below:

1. The three main criteria are: the Physical Message (gestures, eye-contact, etc.); the Visual Message (images, Powerpoint, etc.); and the Story Message (use of English, comprehensibility).

2. Accuracy, pronunciation, fluency and comprehension.

Instructors who reported more holistic grading systems wrote this about their criteria:

3. Communicative ability is secondary to research and the creation of original content.

*Interesting topic. Well prepared. Each member spoke equally.*
Two instructors also reported that they were using peer assessment. Interestingly, when asked to justify their use of presentations, the majority of respondents’ reasons seemed to fall into two distinct categories. The first concerned the importance of gaining confidence in using a second language in front of other people. The second common justification was that presentations are more than just speaking tasks. That is, they require research, planning, writing, the use of gestures, and integration of visual aids as well as speaking. In fact, many instructors using presentations-type tasks cited both of these as influencing factors for their assessment design.

**Speaking Tests**

Of the instructors surveyed, eleven claimed to use speaking tests. About half identified them as their sole methods of assessing speaking, while others used them in conjunction with speaking tasks such as presentations.

The shortest speaking test, in which students gave a prepared speech to the teacher individually, lasted ninety seconds. Taking into account individual, paired and group speaking tests, the majority of instructors seemed to allocate about five minutes per student.

In addition, the style of tests instructors reported as using could also be grouped into two distinct categories: those where the students took the test individually and those where they took it in pairs or groups. Seven instructors assessed students individually, while four reported assessing students in groups and pairs. A summary of both test and rating systems is presented in Table 5.

<table>
<thead>
<tr>
<th>Test Style and Rating Scales</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Test / Analytic Rating Scale</td>
<td>4</td>
</tr>
<tr>
<td>Individual Test / Holistic Rating Scale</td>
<td>3</td>
</tr>
<tr>
<td>Pair or Group Test / Analytic Rating Scale</td>
<td>2</td>
</tr>
<tr>
<td>Pair or Group Test / Holistic Rating Scale</td>
<td>2</td>
</tr>
</tbody>
</table>
While speaking tests were the single most commonly used form of speaking assessment, over half of the teachers did not use them. Their reasons for not using speaking tests are presented in Table 6.

<table>
<thead>
<tr>
<th>Reasons for not Using Speaking Tests</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>It takes too much class time.</td>
<td>4</td>
</tr>
<tr>
<td>It takes too much effort.</td>
<td>0</td>
</tr>
<tr>
<td>The ability of my students is too low so there would be no point.</td>
<td>5</td>
</tr>
<tr>
<td>I think it is inauthentic or unnatural.</td>
<td>5</td>
</tr>
<tr>
<td>I used to do it but I stopped because it is too much trouble.</td>
<td>2</td>
</tr>
<tr>
<td>I would like to but I don't know a good/efficient way of conducting speaking tests.</td>
<td>2</td>
</tr>
</tbody>
</table>

Instructors’ Attitudes about Speaking Assessment

The issue of objectivity in any form of assessment is a complex issue. No matter the type of assessment, whether it is a presentation, a speaking test or observation of in-class activities, and regardless of the grading rubric employed, it seems like it is impossible to eliminate subjectivity from the equation when it comes to speaking assessment. Furthermore, when even comparatively highly regulated and resourced speaking tests such as the Occupational English Test in Australia have inconsistencies (Iwashita and Grove 2003), with our limited time and resources, a truly objective assessment of speaking is an exceedingly difficult goal.

CELE Instructors were asked, “Do you believe your speaking assessment method is objective and if so, how do you attempt to maintain a degree of objectivity?”

Answers to the first half of the question have been divided into four categories: “undecided” “no” “yes,” and “qualified yes”.

<table>
<thead>
<tr>
<th>Do you believe your speaking assessment method is objective?</th>
<th>Undecided</th>
<th>No</th>
<th>Yes</th>
<th>Qualified Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>
The last category is for answers that were generally affirmative but expressed some reservations such as:

1. *I think it's as objective as is possible with something that is so variable and subjective as human interaction.*

2. *Yes and no. I use rubrics with clear parameters to help at least.*

Of the instructors who responded with answers interpreted as “yes” or “qualified yes,” most made reference to having fixed criteria for grading to assist with objectivity and some made reference to communicating those criteria to students before assessment.

Of the instructors who responded that their method of assessment was not objective, two expressed dissatisfaction with their methods of speaking assessment, considering it as something that they were struggling with or hoped to improve. On the other hand, three instructors conceded that their methods of assessment were deliberately or necessarily subjective and suggested that objectivity was illusionary or constrictive of speaking assessment.

**Conclusion**

The survey discussed in this essay sought to gather information about the state of speaking assessment in Freshman English classes at Asia University. Responses were relatively honest and provided information about the various methods of speaking assessment from the perspectives of instructors.

In reference to the first research question, “Which methods of assessing speaking are being used by teachers at the university?” the answer is that instructors are using a range of speaking assessment methods including ongoing assessment, presentations and speaking tests. Instructors provided rationales for their chosen methods while also acknowledging some reservations as to their perceived effectiveness, accuracy or ease of use.

Regarding the second research question, whether speaking ability and achievement is under-represented in our methods of assessment, the answer seems to be, “No.” The majority of
instructors, in accordance with the significant amount of class time devoted to speaking skills, had planned assessment tasks and grade points devoted to speaking assessment. Many instructors also indicated that a portion of participation points was also allocated for in-class speaking activities.

The design of speaking assessment can be complicated and its implementation can also be time consuming. Overall, the results of the survey indicate that instructors at Asia University are putting significant thought into the assessment and improvement of their students’ speaking abilities. On this note, and upon reading such findings, I believe instructors will continue to refine and improve their chosen methods of assessing speaking and exchange ideas with other instructors.
References


