

1. Introduction

There is a vast difference between the Ministry of Education, Culture, Sports, Science, and Technology's (MEXT) official language policy and what takes place in actual lessons. The former, through a belief in the importance of English for Japan's future, espouses a reform of English Language Teaching (ELT), the result of which will place Japan at the vanguard of proficiency in Asia. In other words, training learners to not just merely understanding the language on the page, but to be able to harness it for communication purposes i.e. actually be able to speak the language. ([1]MEXT 2014). In high school classrooms, however, there is little time to practice spoken English to achieve communicative competence due to the pressure placed on teachers to focus on university entrance examination preparation, most likely with the Grammar-Translation method notoriously devoid of any communicative framework whatsoever. In fact, such a teaching methodology reinforces an over attention to detail by championing the idea of there being only one correct answer to situate full understanding by the reader. From this situation, naturally, a fear of making mistakes is the norm which closes down or at best severely hinders spoken communication. Of course, perfect, error-free spoken utterances are rarely required in oral communication, as native-speaker language mistakes can testify. As a way of avoiding these mistakes, Japanese learners of English will choose to remain silent. This can be termed a negative washback, where the negative influence of test preparation damages communicative competence. Talandis [2] (2017) however, believes washback can also be positive. This positivity can stem from two factors, one which influences the teacher directly; the other, the student. For the teacher, a test can provide a clear marker of the effectiveness of the teaching and consequently a useful diagnostic tool to make necessary adjustments to increase uptake; for the student, the test can provide a motivational vehicle, albeit an instrumental one at first. The reality of Japan's test-centric pedagogical culture means that whatever is tested by teachers is taken seriously by students. If some aspect of learning does not have any effect on a grade, students will likely not invest time and effort in it to learn it, (Breux, 2016). As a university oral communication teacher in Japan, armed with this knowledge, it should be possible to create classes which harmonize course and test objectives.

At the time of writing there were no standardized uniform objectives for the oral communication courses at Kobe Kaisei College, as both teacher preferences for dialogic and monologic approaches were known to exist. This teacher/researcher, before utilizing the oral testing approach outlined in Talandis's [2](2017) book, had decided to practice communicative English from a dialogic perspective, as a monologic, speech-based approach is essentially written English read aloud, and obviously makes the teaching of conversation strategies redundant. It is hoped that over time, after further experimentation, adaption of techniques, and peer acceptance, a system can be implemented for use by all oral communication teachers in the department.

This paper details the application of some of the main ideas in “How to test speaking skills in Japan” ([2]Talandis, 2017) as a means of providing more effective EFL oral communication lessons on a Japanese university EFL oral communication course. Testing theories and approaches will be briefly summarized followed by their selected application made for the teaching context in question and the rationale behind it.

2. Types of tests

It is important when talking about tests to clarify the two main types: Norm-referenced and Criterion-referenced. Norm-referenced are standardized tests such as TOEIC or IELTS and are instruments for dividing students into groups based on their proficiency very often for admission to an educational institution or placement on a course but not related to the learning content of the course. Criterion-referenced tests, however, suit the classroom better because they practice and review course material providing positive washback in the form of performance feedback which motivates learners to study further. In the Japanese EFL university teaching context in question, a criterion-referenced approach would be most beneficial.

3. Key principles for evaluating tests

Tests can be evaluated by analyzing them from the standpoint of their reliability, validity, and practicality.

Reliability concerns itself with whether tests are administered fairly and consistently. In practical terms this means students can follow clear test directions in a serious manner on a test they know the material and marking scheme of, and the results of which would not differ if administered by another administrator/teacher.

Validity focuses on whether the test actually does test what it is supposed to test: for a speaking test, does it actually measure speaking, obviously with a typical multiple choice test this would not be the case. It is precisely this mistaken choice of instrument that would lead to negative backwash.

Test practicality must consider the ease by which the test is prepared and marked by teachers, interpreted by students and fits into the parameters of both available budgets and time. Language assessment needs to be efficient in order to quickly gain the necessary information regarding students' progress and any remedial action that may be necessary.

There is no perfect combination of the three and instead the administrator/teacher must surrender to the reality of the constant opposing force of each one in relation to the other two. Nation [3](2013) however advises placing reliability and validity over practicality as far as possible without compromising the latter unnecessarily.

4. Positive washback

As previously discussed, with ELT in Japan there is a dichotomy that exists between official government linguistic goals and an actual teaching context which very often works contrary to them. This results in learners, who despite having six years of English education, are unable to communicate in spoken English. English is primarily a subject for norm-referenced testing for

university entrance examinations which influence how English is taught and learned in schools resulting in negative washback with the knock-on effect of lowering student motivation. By accepting the reality that “what gets tested gets done” ([4]Breaux, 2016), it is possible to use the system against itself. By implementing a criterion-referenced approach, students can understand that the material studied in class directly prepares them for the test and future authentic opportunities to use English.

5. Marking and administering speaking tests

5.1 Holistic or analytical approach

After deciding to adopt the practice of oral testing it is then necessary to opt for either a holistic or an analytical approach. The former provides a single score based on an overall impression of a student’s test performance, which is useful if there are a lot of students to test and with high frequency, for example, every lesson. The downside to the single score is the lack of any diagnostic data providing specific information about progress in the student’s speaking ability, with inconsistency playing a key role in the lack of effectiveness of a reliable score due to different teachers having different personal criteria when it comes to marking. With an analytical approach student test performance can be broken down into individual constructs such as fluency, grammar, interaction and pronunciation with each construct having its own scale, a final score being the total of the individual scores on each scale. In this way, the tests are also easier to give because only one construct is being focused on at any one time leading to higher consistency and reliability. Furthermore, constructs could be weighted, for example awarding the skill of fluency with a higher weighting than grammar.

The class sizes in question ranged from 12 to 17 students with a range of both language and motivation levels. Talandis[2] (2017) suggests more frequent testing for students with low maturity and erratic study habits, to focus productivity and provide a clear reason to practice English. Furthermore, higher testing frequency has a positive impact on overall test reliability and validity: the more data generated, the more accurate the final grade will be. As the students in question had a wide range of ability and motivation albeit erring on the low side, and it was hoped that in future a standardized testing framework could be extended for use by other teachers, it was decided to adopt an analytical approach. Furthermore, Talandis[2] (2017) mentions the potential of testing for teacher development by enabling teachers to increase their knowledge about the workings of a conversation and thereby directly influencing the course and test content.

5.2 Identifying and defining evaluation criteria

An advantage of the analytical approach is the amount of diagnostic data created which can influence positive washback. The constructs chosen for assessment focus student attention which means more time and effort is invested in improving them as inclusion in the test is guaranteed. The positive washback provides the teacher with valuable data to inform the teaching process for subsequent lessons.

With the decision made to adopt an analytical approach it is then necessary to identify and define the criteria i.e. aspects of speaking, that are to be evaluated. In the teaching context in question, no specific, standardized official course objectives existed for use by teachers in the faculty. The teacher, this researcher, through extensive experience with teaching Japanese EFL students who had been previously exposed to a grammar-translation pedagogy in junior and senior high school, decided to prioritize the constructs of fluency, interaction, pronunciation and grammar in that order of importance. Japanese EFL learners are overly sensitive to grammatical accuracy which is the main reason why they find speaking with any degree of speed so difficult: before an utterance is made it is subjected to intense cognitive checking and rehearsing. In order to attempt to overcome this, fluency was made the highest priority. Fluency has various definitions. In the context of the oral communication class it was taken to mean speech which had a smooth pace with very few pauses or hesitation. Interaction was deemed necessary because Japanese speakers of English have a tendency to lapse into silence when they are not actually speaking due partly to cultural norms to show respect to the speaker, but also due to a lack of instruction. Pronunciation was chosen for evaluation because of Japanese English speakers' tendency to be subject to L2 interference from Japanese being a syllable-timed language and the katakana alphabet pronunciation of many English words. Finally, the construct of grammar was included to ensure that comprehension of the spoken message was maintained and not distorted through error.

5.3 Adapting descriptors

To address the principle of Reliability it is essential that students understand what constitutes a high or a low mark by way of descriptors. In order to write simple and direct descriptors for the constructs of Fluency, Interaction, Pronunciation and Grammar, Talandis[2] (2017) advises using Weir's [5](2005) constructs and adapting them to ensure full student comprehension. The descriptors below range from two or three in number depending on the construct being graded, with the higher number indicating a higher mark. Furthermore, the higher levels are made to be challenging but not impossible to achieve which is important for maintaining motivation. The following four descriptors were adapted from Weir [5](2005):

Fluency

- 1 You hesitate a lot and speak slowly with many pauses.
- 2 You generally keep the pace of the conversation but sometimes hesitate.
- 3 You speak smoothly and quite quickly.

Interaction

- 1 You are almost silent when your partner speaks.
- 2 You sometimes react to your partner when she speaks.
- 3 You react a lot to your partner when she speaks and participate in the conversation.

Pronunciation

- 1 It is often difficult to hear and understand what you are saying.
- 2 It is usually easy to hear and understand what you are saying.

Grammar

- 1 It is often difficult to understand you clearly because of your grammar mistakes.
- 2 It is easy to understand you because your grammar is usually clear.

5.4 Scoring each construct

It was decided to give weighted scoring to the four constructs of Fluency, Interaction, Pronunciation and Grammar to reflect the teacher's priorities with a total score of 10, which is easy to calculate percentages.

Construct	Out of 10 points
Fluency	3 (30%)
Interaction	3 (30%)
Pronunciation	2 (20%)
Grammar	2 (20%)

5.5 Marking sheet

As students were usually assessed in pairs the following marking sheet was designed so that the name of the students could be quickly added depending on each individual test's pairings. As the speaking test proceeded it was easy to simply circle a number that corresponded to the perceived level of output for each of the four constructs.

Student:				Student:							
Fl. 1	2	3	Int. 1	2	3	Fl. 1	2	3	Int. 1	2	3
Gram. 1	2		Pron. 1	2		Gram. 1	2		Pron. 1	2	

5.6 Test location

The speaking tests were carried out at the back of the regular classroom with the whole class present. Although there was some embarrassment in the first test, this was not apparent on future occasions with confidence growing markedly through the semester.

5.7 Assigning test partners

Testing was carried out every two weeks: in the class preceding the test, the conversation to be tested was taught and practiced in pairs. On the day of the test, new partners were assigned by way of the regular weekly change to the seating plan (name cards randomly placed on desks). Additional changes were made, as necessary, to cope with absenteeism and any resulting odd number of students. No effort was made to match abilities or to have stronger students paired with weaker ones. The whole pair selection process was random, although ultimately the teacher could influence

seating arrangements if it were considered counterproductive to have certain individual students working together. This scenario, however, never occurred.

6. The test design adopted at Kobe Kaisei College

The assigned textbook for the oral communication classes was the Cutting Edge Third Edition series published by Pearson. Four units were assigned for each semester, with each unit being given a completion time frame of approximately four weeks, and containing two testing opportunities. The second test of each unit usually used the “Task” section at the end of the unit which was a pair based activity requiring learners to complete a discussion based task which eventually reached an outcome. Prior to the actual task, students had been exposed to a similar task modeled as a listening exercise. This listening text also served as a reading aloud text given as weekly homework, and as a listening cloze test to accompany the speaking test on the testing day.

The following test example came from Cutting Edge Elementary, Third Edition, Unit 9. Students have to discuss which Japanese souvenir from a list they had brainstormed in an earlier exercise would be appropriate for each member of a British family (mother, father, daughter and son). The task in the textbook is worded as follows:

1. Think of a souvenir from your country for each person in the Taylor family. Use some or all of the ideas below. Ask your teacher for any words/phrases you need.

- something to eat or drink
- a doll or figure of a famous person
- an item of clothing
- a model of a famous building or place
- an ornament or something useful for the house
- a music CD or DVD

2. Work in pairs. If you and your partner are from the same country, decide on the best souvenir from your country for each person in the Taylor family. If you are from different countries tell your partner about the best souvenir from your country for each person in the family.

A listening cloze activity highlighted the language the students would need to construct their own conversations. This was further consolidated by consulting the audio script and Useful Language ideas in the textbook on the Task page to construct a framework conversation. To help guide less able students and for others to check their own work, the sample below was provided. It was made clear that it was only one possibility and their own elaboration was encouraged.

6.1 Sample conversation for souvenir buying

(use other phrases from the Useful Language Box on p.85 and the Audioscript on p.171)

A: So, let's start with (Bob/Any/Josh/Lorna).

B: Have you got any ideas for him/her?

A: He's/She's an adult/just ____ years old, so how about (category).

B: Yeah, OK, or (another category)

A: I think (first category) is better, maybe, because...

B: Yes, (boys/girls/men/women) love/ really like (superheroes/dolls/etc).

A: Yeah, so a typical Japanese (superhero/doll/etc) would be (name of superhero/doll/etc).

B: OK. Good idea. But, I think (other superhero/doll/etc) is better because (Reason: it's small and easy to carry).

A: Right. I see. So let's get that for (name).

B: Sure. So... What about (another person)? Any ideas for him/her?

A: Well, he /she would probably like...

Due to time constraints, the test would only include a discussion about two people from the Taylor family. In pairs, students practiced the conversation in class numerous times and were encouraged to try and memorize key phrases and consolidate for homework. The following week was the speaking test and after getting a new partner via the seating plan, were allowed time to practice deciding which two family members they would discuss and the ideas for souvenirs they would suggest for each.

The order in which the students were to be tested was done by drawing lots, although any student pairs containing a member who had been absent from the previous week were offered the last testing slot. Approximately 10 to 15 minutes were allowed for practice before the test began.

Following completion of the test, students were encouraged to look over the listening test text to review and prepare, or complete related grammar exercises from the workbook. Immediately after each pair had been tested, brief oral feedback was given by the teacher (tester) to each pair giving reasons why the particular score for each construct was given.

Other tests were produced by using questions from the Units, often originally included to provide the relevant grammar practice for the unit in question. These questions were then practiced after further input for conversation strategies was included. Such strategies included: talking for a longer turn; expanding the topic; and showing interest. Typically, students would select two items each and use each as a way to start a conversation. This had been practiced extensively in the week prior to the test with special attention being given to a conversation strategy. The following is input provided to show interest by summarizing and was used to focus on how interaction was to be marked.

6.2 Showing interest by summarizing

In the following conversation, look at how “B” interacts by summarizing what “A” says. “A” talks about David Beckham.

A: So I’m going to talk about David Beckham.

B: David Beckham. OK.

A: Well, David Beckham is an English footballer who has played for Manchester United and Real Madrid, as well as representing his country 100 times.

B: He’s really experienced isn’t he.

A: He married Spice Girl Victoria Beckham and has become a worldwide celebrity, an advertising brand and a fashion icon.

B: More than just football.

A: Beckham was born in London in 1975.

B: So he’s... 44.

A: Yes. His parents were crazy Manchester United supporters. His talent was obvious from an early age and he signed with Manchester United on his fourteenth birthday.

B: For a long time.

A: He won many trophies with Manchester United, including the Champions League, and won a league title with Real Madrid. He also captained his club and country.

B: A great player.

A: Beckham has many interests off the soccer pitch and often appears in newspapers, especially concerning his marriage and children.

B: Newspapers really love scandals.

A: He lives near Tom Cruise and the two are best friends.

B: Wow! Movie star friends!

7. Future adaptations and improvements

It is hoped to improve the feedback part of the testing process by producing a feedback document such as the one below which would be filled out in duplicate with the marking sheet with the addition of a simple comment in English to suggest ways of improving future output to increase the score.

Name:	Date:	Test:
<p>Fluency</p> <p>1 You hesitate a lot and speak slowly with many pauses.</p> <p>2 You generally keep the pace of the conversation but sometimes hesitate.</p> <p>3 You speak smoothly and quite quickly.</p>		<p>Interaction</p> <p>1 You are almost silent when your partner speaks.</p> <p>2 You sometimes react to your partner when she speaks.</p> <p>3 You react a lot to your partner when she speaks and participate in the conversation.</p>
<p>Pronunciation</p> <p>1 It is often difficult to hear and understand what you are saying.</p> <p>2 It is usually easy to hear and understand what you are saying.</p>		<p>Grammar</p> <p>1 It is often difficult to understand you clearly because of your grammar mistakes.</p> <p>2 It is easy to understand you because your grammar is usually clear.</p>
Score out of 10 =		
Comment:		

8. Conclusion

After implementing the oral testing theories and approaches outlined in Talandis, (2017), on a regular basis in class, their validation can be provided due to the clear improvement in student motivation and class participation noted by this teacher/researcher.

However, before further thought can be given to any standardization of oral testing, oral communication course objectives in the university (Kobe Kaisei College) should be uniform. It is hoped that as a result of this paper's investigation, the implementation of necessary action to stimulate accepted uniformity of course objectives provided by teachers, will provide the framework for the future implementation of successful standardized oral testing. Before such time, the grading of students across the department will remain inconsistent.

9. References

- [1] MEXT. (2014, September 26). *Report on the future improvement and enhancement of English education (outline): Five recommendations on the English education reform plan responding to rapid globalization*. Retrieved from <http://www.mext.go.jp/en/news/topics/detail/1372625.htm>
- [2] Talandis, Jr. J. (2017), *How to Test Speaking Skills in Japan, A Quick-Start Guide*. Kyoto: Alma Publishing
- [3] Nation, P. (2013). *What should every EFL teacher know?* Seoul: Compass Publishing
- [4] Breaux, G. (2016). *How to teach and test English conversation in Japan*. USA: Conversation-based Learning
- [5] Weir, C. (2005). *Language testing and validation: An evidence-based approach*. New York: Palgrave MacMillan