

# Why do oral testing?

By David Evans

The problem of Japanese students being unable to speak English is well known and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has branded the English proficiency levels of Japanese learners 'inadequate'. Consequently, last summer the ministry released its 'English strategy initiative', which aims for 18 year-old school leavers to be 'able to hold everyday conversations in English'. From 2006 public universities should test applicants on their listening ability (McCurry, 2003). Some public universities have already included such an exam, but it is yet to feature in the 'Center Exam'. Listening comprehension should go some way to improving students' comprehension skills, but whether this will automatically lead to an improvement in speaking is yet to be seen. New oral communication classes should help, but if they are not reinforced with a meaningful speaking test there is a risk that they will not be taken seriously.

## Why don't students speak well?

One reason for poor speaking ability is that the students have simply not had the opportunity or reason to practice this skill. English has been taught as an academic subject, as something to know about rather than as a tool to use. The problem is not unique to Japan. For example, Richards (1993) suggests that the inability to speak a foreign language is not dissimilar to American students' inability to speak French or German after many years of study. It has been presumed that if students 'know' how a language works they should be able to speak it. However, being able to recognise a grammatical pattern -- typically presented in examinations -- does not ensure that the student will be able to produce it in conversation.

No one knows exactly what the best way to learn a language is. Theories of second language acquisition are adopted and discarded at quite a pace, and the only consistent theme is that all are found wanting. If we wish to improve any skill, such as a sport or musical instrument, we need to spend many hours in practising it. Learning a foreign language would seem to be no different. Typically, students in Japan learn about English rather than learning how to use it. This is like learning how to drive a car by reading the manual, but without getting in to drive.

At present, the content of English classes is mainly designed to help students succeed in university entrance examinations. Introducing oral testing would change the focus of English classes to preparing students for a test of their ability to use the language, rather than a receptive knowledge test.

## Why do oral testing?

The notion of 'backwash' or 'washback', as it is sometimes referred to, is defined by Hughes (1989:1) as 'the effect of testing on teaching and learning'. If we assume

that the majority of students study to be successful in examinations at school, they would be wise to devote their energies to studying for what will be tested in their exams. Equally, teachers will want to prepare their students to do as well as possible so what both students and teachers do will be determined largely by the content of the test.

Backwash can be either positive or negative depending on whether the test encourages or discourages teaching and learning behaviours that one wishes to encourage. It would seem that if the goal of English language education at school is for students to be able to speak the language, then the current university entrance exam system and 'high stakes' tests, such as TOEIC and TOEFL have little positive impact on encouraging speaking. As practising speaking is not necessary for success in these exams, there is no need for students or teachers to devote any time to it. Speaking all too often becomes an optional extra, rather than a central pillar of the syllabus.

If, however, the exam encourages students to achieve the goals of the curriculum then the backwash effect could be said to be positive. So, if conversation is the aim, the exam should encourage students and teachers to spend time on both speaking and listening.

The current exam format for English in the Centre Exam does not require students to speak English. What it tests, the majority of students are able to do. Helgesen (1993) and Richards (1993) both comment that Japanese students are capable of doing what they studied; that is, translation of grammatical structures, analysing, reading comprehension, vocabulary recognition and such like. However, being able to do such activities does not equate to being able to speak. There is also the question whether it is really possible to measure speaking ability in anything other than a 'direct test', i.e. a speaking test.

The introduction of a listening exam in addition to the 'paper' test in the Center Exam from 2006 would seem to suggest an acknowledgement of the notion of backwash, as does the increasing inclusion of direct speech in reading comprehension passages. Whilst both of these measures may have some positive 'backwash' effect, the bolder step of introducing an oral test into the university entrance exams would undeniably change the priorities of both teachers and students.

Indirect testing, which for speaking would mean any test that is not a spoken one, is one that 'attempts to measure the abilities which underlie the skills in which we are interested' (Hughes 1989:15), but whether such tests can test the ability to produce language rather than merely to recognise it, is increasingly being questioned (see Heaton (1988), McNamara (2000) and Shohamy (2001)). Hughes believes that the abilities we wish to develop should be tested, so that "if you want to encourage oral ability you should test oral ability" (1989:49). Heaton asserts that if we

test indirectly, “we are removing an incentive for students to practice in the way we want them” (1988:45).

Also fuelling concerns that such tests cannot measure speaking ability are plans to introduce a speaking component into the TOEFL examination, reflecting scepticism that a student with a score of 550 can communicate sufficiently to study at a foreign university. Likewise, apart from the highest scores in TOEIC (over 800 or so), student scores at lower levels do not seem to reflect the students’ speaking ability. For an interesting critique of the extent to which TOEIC measures communicative competence, see Cunningham (2002). In my own experience, returnees, who have been able to hold a conversation in English with ease, have complained that they are ‘bottom’ in English at high school. It seems very likely that most ‘high stakes’ tests such as those mentioned are not reliable guides to how able a student is in speaking.

A direct test of speaking (an oral exam) as opposed to an indirect test (such as a multiple-choice test) could ensure that the positive backwash effect would lead to more classroom time being spent on developing the skills necessary to communicate.

### What is speaking?

David Brazil (1995) makes the important point that speaking takes place in ‘real time’, meaning that there is great pressure to both comprehend and produce language immediately. Referring to native speakers, Brazil notes, ‘our ability to put together what we want to say may not always be equal to the pressure to keep up with ourselves, so to speak, in the delivery of our message’ (p.11). This problem is particularly acute for non-native speakers, and both comprehending and creating language in real time can only be tested in the spoken mode.

There are other aspects of conversation that cannot be properly tested in a written test. Cornbleet and Carter (2001) also point out other features of conversation. Conversation is ‘face to face’ which means that we change what we say in response to the immediate feedback we get from the other person’s face, for example, knowing when to interrupt and take our turn in the dialogue, or knowing if the question we have asked is something that our partner wants to talk about. Such features cannot be tested unless in a conversation.

Conversation also has a ‘purpose’, even if that purpose is ‘just’ to maintain a relationship. We do not speak in the real world to display our knowledge of grammatical accuracy, and as Brazil states, “speech is characteristically used in pursuit of purpose...the practice of inventing a sentence...is a practice of the sentence grammarian, not the user”. We do not speak in sentences but in utterances, which do not have to be grammatically correct nor concise. Speaking ‘correctly’ is not the goal of conversation and native speakers do make mistakes, and these mistakes do not usually hinder the conversation. However, it seems that for non-native speakers we demand more precise English than we do for those conversing in their mother tongue.

In indirect testing however, recognition of grammatical



accuracy would seem to be of disproportionate importance, simply because it can be measured, whilst other aspects of communication cannot. This has a further negative impact on students in that when they do attempt to speak English, their speech can be painfully hesitant; as they are loath to make mistakes.

It also inhibits risk-taking with the language. Making mistakes is an integral part of learning but multiple-choice tests reinforce the notion that English is right or wrong above all else, and this slows students’ progress in speech.

### The problems with oral testing

There are some understandable reasons for the reluctance to use oral testing for ‘high stakes’ testing (such as the central university entrance exam in Japan). One is the difficulty in defining exactly what it is that one wants to test in speaking. The notion of ‘communicative competence’ as defined by Hymes (1972), then refined by Canale and Swain (1980) and later by Bachmann (1990) (summarised in H. Douglas Brown, 1994), has given us an insight into what it means to be able to communicate. It is clear that ‘grammatical competence’ is but one necessary component of communicating, but perhaps because it is the ‘easiest’ to test objectively it remains the centrepiece of most tests. Other aspects such as ‘strategic competence’, meaning, among other things, the ability to sustain a conversation using such tools as guessing and hesitation, are very difficult to assess in an analytic manner. However, it is easy to recognise how important such aspects are if a conversation is not to die a premature death, and an examiner should be able to detect the essence of such a skill.

Other competencies such as ‘sociocultural competence’, which includes interaction, might be difficult to test in certain formats of oral testing so this too is a potential prob-

lem. Rating scales with sub-divisions such as fluency, complexity, appropriateness and so forth can give an outline for how to grade, and by choosing those aspects of conversation that one wants to encourage, students and teachers will gear their efforts towards accomplishing them.

The second main problem with oral testing is the issue of reliability. The notion of test reliability requires that if the same student were to take the same test on two different occasions within a limited space of time, he or she should gain a score that is comparable to the first occasion. If the score is wildly different, then the test is unreliable. We would not expect the student to get exactly the same score, as humans are not machines, but neither would we expect the student to get, for example, 80% on one occasion and 23% on the other. Grading of oral tests is impressionistic and therefore reliability cannot be guaranteed. It is difficult enough for one grader to be consistent over the course of a day's examinations, and this problem of variety in scores is multiplied with an increased number of graders. Steps can be taken in the training of these examiners to make them as consistent as possible, but it would be a folly to assume that one standard can be maintained.

This problem of the possibility of error however, currently exists within multiple-choice tests. There is a one in four chance that any answer may have been guessed, yet this is deemed to be satisfactory. The percentage of a candidate's score that can be attributed to guessing is unknown in the multiple-choice format, yet the ability to make lucky guesses must be determining the academic future of students every year.

## Conclusion

If the Ministry of Education is serious about the need for Japanese students to be able to maintain a conversation after graduating from high school, then employing an oral test would change the perceptions of both teachers and students as to what is important. Currently teachers are doing what is best for their students' futures, by preparing them to do as well as possible in the Center Exam and other university entrance exams. It is because of what these exams test and the nature of backwash that little attention is devoted to the skill of speaking. Introducing an oral test would oblige both teachers and students to change their approach to English. Obviously, such a change would be hard for all concerned and there would be the risk of some errors in such a test, but the benefits would almost certainly outweigh the negatives.

Other issues would also need to be considered. Most importantly, is the question of how speaking is best taught in the Japanese environment and what additional training would be necessary to ensure that teachers feel comfortable with this different style of teaching. One positive impact of backwash on students however, is that once they know they will take oral tests, speaking activities cease to be a light-hearted diversion and they will apply themselves to these new activities. It can be embarrassing and seemingly pointless for one high school student to practise speaking English with another, but knowing that this skill is to be

tested makes it a far more worthwhile undertaking for the student.

Another important consideration would be the style of test adopted; there are many different formats, each with its own strengths and weaknesses, and each with its own backwash effects. For example, the individual interview is unlikely to replicate conversation as research has shown (Kormos 1999) that students do not feel able to ask an examiner questions. An alternative is to have students tested in pairs or groups, which can circumnavigate this problem. However, in groups there is the risk that a poor group might negatively impact on an individual. The sheer scale of such an endeavour, such as training the necessary examiners and scheduling the tests is also colossal. However, despite the difficulties involved in using oral testing its use as a meaningful university exam would lead to fundamental changes in the way that both teachers and high school students regard English, resulting in more time and attention being given to spoken communication. This, in turn, should help realise the goals of the government concerning the desired levels of students' spoken English.

## References

- Bachman, L. (1990) *Fundamental Considerations in Language Testing*. New York: Oxford University Press
- Brazil, D. (1995) *A Grammar of Speech* Oxford: Oxford University Press.
- Brown, H. Douglas, (1994) *Principles of Language Learning and Teaching*. (pp226-230) New Jersey: Prentice Hall Regents.
- Canale, M & Swain, M (1980) "Theoretical bases of communicative approaches to second language teaching and testing" *Applied Linguistics* 1: 1-47 Oxford University Press
- Cornbleet, S. & Carter, R. (2001) *The Language of Speech and Writing*. London: Routledge.
- Cunningham, C. (2002) *The TOEIC Test and Communicative Competence* available online at: [www.cels.bham.ac.uk/resources/essays/Cunndiss.pdf](http://www.cels.bham.ac.uk/resources/essays/Cunndiss.pdf)
- Heaton, J. (1988) *Writing English Language Tests* New York: Longman.
- Helgesen, M. (1993) "Dismantling a wall of silence. The 'English Conversation' Class." In Wadden, P. *A Handbook for Teaching English at Japanese Colleges and Universities* (pp 37-49). New York: OUP.
- Hughes, A. (1989) *Testing for Language Teachers* Cambridge: Cambridge University Press.
- Hymes, D. (1972) "On Communicative Competence. In Pride and Holmes" 1972 *Sociolinguistics* Harmondsworth. Penguin.
- Kormos, J. (1999) "Stimulating conversations in oral-proficiency assessment: a conversation analysis of role plays and non-scripted interviews in language exams" in *Language Testing* 1999 16 (2) pp 163-188.
- McNamara, T. (2000) *Language Testing* Oxford: Oxford University Press.
- Richards, J. (1993) "Real-World Listening in the Japanese Classroom." In Wadden, P. *A Handbook for Teaching English at Japanese Colleges and Universities* (pp 50-62). New York: OUP.
- Shohamy, E (2001) *The Power of Tests* Harlow: Longman.
- McCurry, J. (2003, January 30) Is Japan ready to start talking? *The Guardian* <http://education.guardian.co.uk/tefl/story/0,5500,884801,00.html>

*David Evans has taught in Japan for 13 years. He currently teaches English at the National College of Nursing, Japan.*