Oral Testing and Self-Assessment - The way forward?

Hyun Taeduck & Andrew E. Finch
Andong National University
April 1998

Abstract

Traditional methods of testing which still prevail in Korea mean that there is very little time for Conversational English in the test-driven secondary classroom, despite introduction of recent advances in language teaching methodology. This paper suggests that the use of English can only be promoted in schools by incorporating an "Oral Test" into the overall testing requirements, and that self assessment by students of those tests can be a valuable additional means of improving oral abilities.

A test was developed focusing on the improvement in spoken English of 1700 Freshman University students over an academic year (64 hours). This was administered and evaluated using established oral-test criteria. Grading was relative, looking at improvement rather than level of achievement, and the Conversation-English course taken by the students was the basis of the test. Results showed that: i) preparation for the test necessitated active spoken participation in lessons; ii) lessons tended to utilize task-based communicative teaching methods; iii) the means became the end - the test was not only a reason for developing oral skills, but also a means of achieving that goal.

I. INTRODUCTION

The issue of oral testing highlights a major problem for educators in Korea, where an official policy to promote English oral skills at all levels of education has to be placed in the context of the traditional methods of testing which still prevail (Lee 1991:47). Thus while secondary level teachers of English might be keen to employ contemporary communicative teaching methodology in their classrooms, they also have to ensure that their students acquire the necessary linguistic 'facts' to be able to answer the grammar-based multiple choice examinations which are universally prescribed, but which tend to be unrelated to the development of spoken English abilities (Lee, 1991:342). The result is that students in general arrive at University or their new place of work with undeveloped oral skills and with a debilitating awareness of this fact, which impedes motivation or further improvement.
Years of traditional translation-based teaching also encourage learning preferences in the students which, though acknowledged by them to be inefficient, are all that they know, and determine their perceptions regarding acceptable teaching and learning styles (Brindley 1984, Horwitz, 1985, 1988). Such learner beliefs are identified by Harris (1997:14): “students' perceptions can clash with the procedural goals of communicative foreign language learning”, and Mantle-Bromley (1995:375): “For many students, their beliefs about the nature of language-learning may constitute a serious impediment that could affect their language-related attitudes and behaviors”, and because of this, more effective but unfamiliar teaching and learning methods can be unwelcome, and the use of activity-based language activities can be seen as a diversion from 'real learning'.

In order to break this self-confirming circle, and to motivate students to develop their oral skills, the situation needs to be addressed in its entirety. Teachers can find more and more interesting methods of teaching the spoken language; they can try to apply these in the classroom, advocating authenticity of materials, relevance of situation, cultural sensitivity, and other factors; they can make the learning environment as conducive to expression and language acquisition as possible. But the fact remains that Korean students are motivated mainly by the National exams they have to pass, and their entire educational experience confirms that this attitude is the 'correct' one.

Having identified the tests as the driving force behind learning in schools, there are good arguments for adjusting those tests to be more communicative and to incorporate evaluation of oral abilities (Lee, 1991:342). However, the researcher needs to proceed with caution in this area, since there are a number of pitfalls: i) we must beware of producing new tests which purport to address the issues involved, but which are in fact similar in design (and therefore in effect) to those which they replace; ii) if we focus on making the course content more conversation-based, but leave the tests as they are, this will be unfair to the students, who will have to study according to one set of parameters during class time, while preparing for a test based on differing principles; iii) if we drop the established test and use only continuous assessment of oral abilities, the course will not be seen as valid by the students, who have been taught that they need to 'see results'.

The approach suggested in this study, and the method adopted by the Language Center of Andong National University, is to design oral tests of communicative competence (Morrow 1981:18) according to the learning goals set for the students and taught in the language program, so that these tests will have a wash-back effect on the courses. If the aim is to promote oral skills in the target language, then it seems reasonable that this ability should be at the heart of the testing, and that any test-driven learning that takes place should be directed towards such a goal.

II. RATIONALE
1. Preliminary

The remit of the Language Center is to provide Conversation English courses for all Freshmen, Sophomore and Junior students in the University, and while testing the effectiveness of such a program is essential, the method of doing this is crucial. The English program employs Task-based teaching as a means of encouraging authentic use of English in the classroom, using an in-house textbook promoting this approach (Finch & Hyun, 1997. The syllabus is contained in the textbook, and the communicative activities in which students engage during lessons prepare them for the final assessment. Such active participation is at the heart of the task-based approach, a fact reflected in the weighting given to continuous assessment of classroom performance (65%) in the final grading of the students.

2. Considerations

Producing and administering tests that assess oral skills has the washback effect of encouraging students to acquire those skills in order to pass the tests. However, once the decision has been made to design oral tests based on the students' needs, there are still problems of which the educator needs to be aware. As West (1988:8) remarks:

- Testing of the spoken language is now established as an essential feature of examinations in Britain, but there are still parts which have not been reached by the communicative beer can. But it is essential that the testing of spoken English should reach these parts.

Thus the oral test must be a true assessment of spoken abilities, rather than an indication of how well a student can read a passage in English, or can produce well-memorized responses that have little meaning for him/her.

3. The Human Approach

There are various factors which are important in designing a true test of communicative effectiveness, and two of these are authenticity and a relaxed atmosphere, for the essence of a communicative act is its day-to-day setting, free from the constraints of having to conform to linguistic codes. If communication can occur despite grammatical mistakes, then those involved will usually overlook them. Only if there is a problem in transferring information or opinions can this be termed an error in the use of the language, and the oral test needs to recognize this.

Underhill (1987) points out other factors that need to be considered when designing an oral test: full local knowledge; a human approach; a suitable balance; and the ability to adapt and improve
the test. He also mentions that it should be designed as a whole (see also Morrow 1981:18) in
direct contrast to the discrete code-based view of language as evidenced in traditional testing
methodology. Furthermore it is vital that the learner be taken into account during this process, for
as Blanche (1988:75) reminds us: “Students need to know what their abilities are, how much
progress they are making and what they can (or cannot yet) do with the skills they have acquired.
Without such knowledge, it would not be easy for them to learn efficiently." It is easy to lose sight
of the fact that tests must serve student needs as well as the teacher needs, and that an intimidating,
artificial construct designed only to provide an acceptable “bell-curve” will not help the testee, and
can even retard progress by demotivating him/her.

III. METHOD

1. Setting the scene

- Language is culture in motion. It is people interacting with people. ... the most effective
  programs will be those that involve the whole learner in the experience of language as a
  network of relations between people, things, and events (Savignon 1983:187)

How do we design a test that mirrors the student's abilities to communicate in such a network?
How do we make a test which is a true indicator of communicative ability and which gives as much
information to the student about his/her language abilities as it does to the examiner? To answer
these questions we need to ask why we are making the test and what are its goals. Who is it for?
What is it meant to test? How will we determine if it is successful? Do we want reliability or
validity? Is norm referencing or criteria referencing appropriate? It was decided at the Language
Center to answer these questions by designing a test that mirrored what had been taught (high face-
validity), that was learner-centered (Nunan 1988:134), and that had “washback" validity (Morrow
1985). It should be seen as working for the students rather than against them; it should offer
constructive help rather than non-directed criticism; it should show them how they were faring on
the road to fluency and point the way forward; and it should be about assessment of individual
abilities and improvement rather than a comparison of proficiency levels.

The success of this test would be judged primarily by its effectiveness in favorably affecting 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

2. The Test

Of the many possibilities open to testers in terms of elicitation techniques for oral testing
(Underhill, 1987:27), the methods adopted in this case were the oral report and inter-learner joint discussion, the latter reflecting the interactive aspect of language mentioned by Savignon (1985).

Two examiners were present for each test - the normal class teacher, and another 'visiting' examiner (who was also an instructor at the Center and known to the students by sight if not by acquaintance), providing both a subjective and an objective assessment of the students. Students were required to complete two stages. In the first stage each student had one minute in which to speak about him/herself, his/her family, hobbies, room, or lifestyle. These topics had been practiced and performed in the lessons during the first semester. In the second stage the group had three minutes to have a conversation about anything they wished. This was intended to promote interactive skills as well as speaking per se, and the groups were assessed as a whole for this task.

The tests were recorded on cassette tapes and the examiners made comments on a pre-prepared mark sheet, and noted the conversational abilities of the students as they performed the two stages. Categories for assessment focused on communicational effectiveness, and were based on the type of 'performance criteria' mentioned by Underhill (1987:96), i.e. 'size' (length of utterances), 'complexity', 'speed', 'flexibility', 'accuracy', 'appropriacy', 'independence', 'repetition', and 'hesitation'. North (1990:96) simplifies these into 'range', 'accuracy', 'delivery', and 'interaction', while Lee (1991:Appendix 6) reflects the Korean situation by adding 'fluency', 'organization' and 'pronunciation' (see Lee 1991:280 for an analysis of some published rating scales for oral testing). Given that the students would be able to access these mark sheets in order to help in their own assessment of their performance, it was decided (by all the examiners) to use a simplified version of Lee's criteria, employing the categories of 'Listening comprehension', 'grammatical Appropriacy', 'Ease of Speech and Fluency', 'Content' and 'Conversation Skills'. Marking criteria were drawn up based on established principles (Lee 1991:293, Nunan 1988:124, Underhill 1987:98), to provide a checklist of proficiency levels.

3. Reliability-Validity

There is various discussion in the literature on the validity (face/content/construct), and the reliability (concurrent/predictive) of oral tests. As Jakobovits argues (1970:95): “the question of what it is to know a language is not yet well understood and consequently the language proficiency tests now available … attempt to measure something that has not been well defined." Thus communicative testing has a problem which is not shared by the psychometric objective testing which sees language as a definable series of structures rather than as a means of communication. Carroll (1981:8) highlights the special role of oral testing and states that language “must be taught and tested according to the specific needs of the learner." If we define those needs under the heading of “communicative competence", then we must be ready to accept “the trade-off between
reliability and validity" (Underhill 1982:17) involved, but we must also not lose sight of the students themselves, and their perceptions of what is happening. As already mentioned, student beliefs and attitudes are a powerful influence on the amount and quality of learning that occurs (Mantle-Bromley 1995:375. Many students for instance, (especially in the Korean situation) feel the need to be assessed by a professional educator, a person who 'knows' more about the learning process then they do, and for them this is part of the face validity of the test.

IV SELF-ASSESSMENT

Self-assessment provides us with an interesting perspective on this issue. If the learners themselves determine what is to be learnt in the classroom, regardless of what the teacher brings into it (Allwright 1984:4), and if their attitudes to learning are so formative, then it seems that we should be giving more attention to these matters, and focusing on ways of improving them. As Mantle-Bromley (1995:383) has it: “If we attend to the affective and cognitive components of students' attitudes, as well as develop defendable pedagogical techniques, we may be able to increase the length of time students commit to language study and their chances of success in it.”

Self-assessment is a way of attending to such attitudes, since it encourages the student to become part of the whole process of language learning, and to be aware of individual progress (Harris 1997:15). This is a large topic in itself, and one still quite young, but it has been shown that “the validity of learner judgements can in fact be quite high” (Oscarson 1989:2), and that “a majority of students find it easier to estimate their purely communicative competence level than their mastery of grammar” (Blanche 1988:75). Oscarson (1989:3) puts forward a “rationale of self-assessment procedures in language learning”:

1. promotion of learning;
2. raised level of awareness;
3. improved goal orientation;
4. expansion of range of assessment;
5. shared assessment burden;
6. beneficial postcourse effects.

Harris advocates the use of self-assessment in the school classroom, stating that it is a “practical tool” that can “make students more active” and can “assist them with the daunting task of learning how to communicate in another language.” (1997:19).

These considerations are gradually being incorporated into the language program at the Language Center, especially since “this kind of self-appraisal would be particularly helpful in the case of
(false) beginners" (Blanche 1988:86), and a large part of the syllabus for the Sophomore students now focuses on reflective learning. At the time that this study took place however, the main self-assessment that occurred was the filling in of a self-assessment questionnaire at the beginning and the end of the semester. Students were also allowed access to their mark sheets and to the cassette tapes of the testing sessions.

V RESULTS

1. Disadvantages

When analyzing the results of this study, the weaknesses inherent in its design have to be acknowledged. Firstly its subjects were 1,700 students who had no previous experience of oral testing, and who therefore could provide no indications achievement levels. Secondly the classes consisted of students grouped according to their parent Departments rather than any English ability, and ranges were wide both inside classes as well as between them. Thirdly it was not possible or desirable to use the tests to construct a comparative grading of the students, since their backgrounds were sufficiently different to make this meaningless. If we compared the student who had recently returned from America (for example) with one who had learnt all his/her English in a local village school, there would be obvious discrepancies.

Improvement in communicative competence was therefore used as the guiding criterion. Given the lack of reliable indicators of initial ability level however, the only yardstick which could be used was the students' own appraisal of their abilities when they started taking the Conversation course. Therefore examiners applied the same rating scales to all students to determine the present level of communicative competence, and then matched these against the students' self-assessment from the beginning of the academic year. Students also filled in this questionnaire a week before the test, and were able to compare their own assessment with that of the examiners.

There were no control groups, either in the University, or in other Colleges, and so it would be meaningless to try to assert that students had made progress in any significant manner that would not have been made had the test or the teaching procedures been different. However, such quantification of results was 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 the students, who arrive with extremely undeveloped practical skills, and contrastingly well developed confidence barriers and anxieties about speaking English. A prime aim of the English program is to address those problems, and to help the students to perceive themselves as successful learners. Anxiety can be a strong inhibitor of performance, especially in oral tests (Macintyre 1995:96), and success in achieving such a goal cannot be quantified in the short term, but must be viewed by looking for
increased motivation and application over the long term.

2. Advantages

Despite these reservations, it was observed that students following the task-based curriculum and taking the communicative oral tests did show improved performance in spoken English compared with those who had studied English in previous years and were still on campus. However subjective such a view may be, the general involvement of students in the language program, and their readiness to engage in conversation with native speakers is noticeable at this point.

Having administered the test, cassette tapes and marking sheets now exist for each student, and further tests will be able to use these to assess improvement more reliably, and to produce individual student learning profiles. Meanwhile, there are some spin-off effects that have been beneficial to students and educators regardless of questions surrounding the issue of oral testing. One of these is the washback effect of the test. Korean students tend to be test-driven (Lee 1991:44), and the requirement of having to 'perform' in English in the test means that students and teachers must concentrate on practicing 'performance' in lesson times. This leads us to a second effect, the realization that skills development is “a gradual, rather than an all-or-nothing process” (Nunan 1988:5), with students beginning to see that if progress is to measured in terms of performance (Harris 1997:15), then leaving study to the last moment is not an effective strategy.

A further beneficial effect of the emphasis on communicative competence in the test is that because every student is required to speak for one minute about his/her personal situation, we now have a student population that can engage in English (if only for a short time) in a conversation on personal matters. When we look at the sort of content that would be needed for likely situations requiring the use English in Andong, we can see that such an ability is indeed appropriate and useful. This goes against the requirement of spontaneity to some extent, but we must remember that much of spoken language is repetition, so that the effort spent by the students in thinking about the content and in rehearsing it constitutes a valuable learning experience.

Finally, the fact that the students do all the talking in the test eliminates much of the spoken-test-anxiety mentioned above, as well as undesirable results which can arise from mismatches of cultural expectations, termed “perplexities of culturally mixed teacher/student pairs” by Hofstede (1986:302).

VI CONCLUSION

This study has attempted to show that the use of oral tests focusing on communicative competence in schools and Universities will have the beneficial 'washback' effect of ensuring that the courses
focus on means of promoting oral skills. If the development of spoken English is an aim for Korean educators, then the incorporation of an oral test into the present testing system is to be recommended. By administering tests which not only assess the level of oral skills, but which assist in the very improvement of those 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 time to grow. By doing this, the student is acquiring beneficial learning habits and the test is therefore fulfilling more than one pedagogical aim.

The oral test should be an extension of class-work, and of the continuous assessment that occurs in each lesson. However, students can be further motivated to learn and to develop their practical skills by involving them in the assessment process and inviting them to monitor their progress over the semester as well as at test time. As Harris points out: “Above all, they can be helped to perceive their own progress and encouraged to see the value of what they are learning.” (1997:19)

REFERENCES


Brindley, G. 1984. 'Needs analysis and objective setting in Adult Migration Programs' Sydney: NSW Adult Migrant Education Service.


Horwitz, E.K. 1985. 'Using student beliefs about language learning and teaching in the foreign language methods course' in Foreign Language Annals, 18, No.4. pp. 333-40


Morrow, K. 1985. 'The evaluation of tests of communicative performance" in Prospect, 1, 2.


Nunan, D. 1996. 'Towards autonomous learning: some theoretical, empirical and practical issues' in


Underhill, N. 1982. 'The great reliability validity trade-off: problems in assessing the production skills", in B. Heaton (ed).
