Reflection on the Quality of an Impromptu Speech as a Component of an EAP Evaluation

Aytaged Sisay[*]

Introduction

English for Academic Purpose (EAP) courses have been in offer in Ethiopian Higher Education. The focus of these courses has changed as the curriculum of the institutions change. In the recent past, there were two courses (College English I and II) on English for Academic Purpose that emphasized on writing, reading, listening, speaking, vocabulary, and grammar. The main objective of the courses has been to help students develop apposite language and academic skills to pursue their mainstream courses in higher education successfully.

The evaluation aspect of the courses, however, used to disregard the speaking skill at least in one higher education institutions (Alemaya University) where I and other four English language instructors used to work before some years ago. One of the reasons for this boiled down to practicality question that emerged from the department's concern to reduce subjectivity in evaluating students' performance. As a result of this concern, the department's preference was to administer similar evaluation instruments across all sections that took the EAP courses. In other words, testing listening, mid and final exams that comprised all the language skills, except speaking, but including vocabulary and grammar used to be prepared centrally and employed to determine students' achievements based on norm referencing approach.

The practice that gave no or little attention to evaluate students' speaking skill which covered a reasonably good portion of the courses was challenged by many of us who were teaching the courses. Consequently, we started

[*] Lecturer and Researcher, Institute of Educational Research, Addis Ababa University

deliberating on how to include speaking skill as a component of the courses' evaluation. We also tried to ponder over the department's concern of reducing subjectivity in awarding students' grades. What came to our mind along this line was to be as similar as possible in employing similar speaking evaluation mechanisms across all the sections. Some of our efforts with regard to this included a decision that impromptu speech could be used as an evaluation format across all the EAP sections and an identification of forty speaking topics on which each student choose a speaking topic on a lot basis to deliver 3 minutes talk. We (the teachers) used this format to determine students' speaking performance of the semester. Besides, the dates in which this evaluation should be conducted by all of the teachers were set by the department to be the last three days before the academic calendar for the semester indicated the final class.

Teachers' class room decisions are very much interpretative. What holds a proper action to address certain thinkings or problems in one context may raise different questions. In this paper, I would like to retrospectively reflect on our evaluation practice and decision. To structure my reflection on the quality of the evaluation instrument (impromptu speech) as a component of an EAP evaluation, I use concepts like reliability, validity and wash back validity.

RELIABILITY

The idea of reliability can be presented in the following manner. For Weir (1988, p. 34), "the concern here is with how far we can depend on the result that a test produces… could the results be produced consistently?" And for Bachman (1990, p 160), it "… is concerned with answering the questions: how much of an individual's test performance is due to measurement errors or to factors other than the language ability we want to measure?" We, therefore, can trust the impromptu speech as a reliable test if it yields accurate, consistent, fair and dependable scores. The oral test under discussion, however, seemed to lack the above preconditions. As I was one of the testers, it is in my memory that all of us were highly focusing on

administering the test and we were not giving a deserving attention to approximately balance the difficulty level of the speaking topics which has an implication on grading students. This, therefore, may imply the possibility of different performances on different topics. Furthermore, the scores awarded as a result of this test were impressionistic based on our individual subjective judgment. In other words, we have not used a common rating scale. Consequently, these differences in the scoring method had led us to have instability in the test scores. This, indeed, was limpidly observed when we met to discuss and decide cut off points to award grades at the end of the semester. What is more, inter and intra reliability were not checked to see whether or not the scores obtained were the students' true scores. Thus, it seems to be cogent to be skeptical about the reliability of the test.

Hughes (1989,p.105) suggests, " …to make the oral test as long as feasible because it is unlikely that much reliable information can be obtained in less that about 15 minutes, while 30 minutes can probably provide all the information necessary for most purposes." However, in the test under focus, each student was expected to deliver 3 minutes talk, which seems to cast a cloud over the reliability of the information obtained, according to Hughes (1988). Furthermore, it is very vital to know that the scores on language tests differ not only as a consequence of the testees' language abilities but also due to the nature of test formats. Factors such as"test takers cognitive and affective characteristics, their real world knowledge and factors such as their age, sex and background" are presumed to be potential sources of test bias (Bachman, 1990). Due to dearth of knowledge in the effect of test techniques, Alderson et al (1995) advice testers to utilize more than one test techniques to evaluate a given ability. We, however, employed only a technique (impromptu speech) to assess students' oral proficiency. Would this imply the less attention paid to the fact that a given student may perform better in one technique than the other regardless of his/her ability? If so, it seems to be a perplexed condition for this test to have procured accurate and consistent information about the testees.

VALIDITY

The impromptu speech used to assess the speaking ability could be considered as a valid testing technique if it measures the traits it needs to measure accurately and if it samples some defined domains representatively (Hughes, 1989 and Cunnings, 2003). However, the effectiveness of the technique under focus was questionable to gratify the above criteria.

As it is said above, the main objective of the course is to help students interact in English for academic purposes successfully. To this end, students in the speaking section of the course were exposed to different speech styles which may enable them meet the academic demands of a university. The first four units of the speaking section of the course college English I (Atkins et al., 1996), for instance, deals with the following sub-skills of speaking.

- Interviewing friends to find information;
- Reporting interview results;
- Agreeing and disagreeing with ideas;
- Comparing views and preferences;
- Giving suggestions;
- Expressing feelings;
- Negotiating on issues
- Delivering public speeches;
- Discussing ideas based on pictures;
- Discussing ideas to compare and improve notes;
- Presenting persuasive speeches;
- Comparing answers;
- Asking questions etc.

As can be seen from the above list, students were expected to engage in different speech styles. Nevertheless, the assessment was made through an impromptu speech which may be amenable for a single speech style. Would this imply that there were some sub-skills of speaking which were under

represented or not represented at all in this test? If so, it is questionable whether or not performance on this test could be a valid indication of the students' over all oral proficiency, which included additional sub-skills. Moreover, the impromptu speech alone may not be a representative sample of the population of oral tasks that a College English instructor would expect his/her students to be able to perform. To this effect, this testing task might be suspected whether or not it helped us to elicit the expected behaviors which truly represent the students speaking ability. Besides, it is also difficult to justify the quality of the impromptu speech from linguistic and educational points of view that a single speech style includes the use of different linguistic and socio-linguistic features in oral discourse which are elements of the language proficiency that students are expected to acquire at the end of the course.

Furthermore, the oral test under discussion seemed to have undesirable appearance to many of the students. For example, as the topics of the speech were divided on a lot basis, most of the students had complained about the unfairness of the oral test. Consequently, many of the students seemed to expect or at least to hope for different oral performances if they were given another speaking topic. Similarly, the majority of them seemed to feel that they were exposed to a test technique which they had never experienced in the classroom. As they were required to deliver the speech in front of the class, a considerable number of students felt that their performances were affected by stage fright. In like manner, some of them had considered the preparation time given to deliver the speech to be not sufficient. To this end, one may question whether or not this test had provided reliable information about the testees. One may also continue doubting the face validity of the test for it did not seem to be in a position to evoke positive attitudes on the part of the testees.

Wash Back Validity

Testing can have a positive or negative effect on teaching (Alderson et al., 1995). Thus, testers, according to Hughes (1989) can sample wide language

areas, test directly what they want to test and make test criterion referenced to achieve beneficial wash back. The speaking test under discussion, however, seemed to have some drawbacks to bring about positive wash back effect on the teaching-learning process. As it is pointed out above, the speaking ability of each student was evaluated on a single oral performance test. This evidently divulges that a student had only one chance to mirror his/her oral competence during a semester. To this end, the majority of the students may question their oral interactions in their speaking sessions as obtaining no or petty contribution to their final speaking scores. This may be taken as one of the factors that make students reluctant to participate in group discussions, a very common problem in most of the English language courses in the context under discussion.

What is more, since the speaking test topics were determined on a lot basis, a considerable number of testees might have thought that success in speaking evaluation was not only determined by the level of proficiency students procured but also by lack which again seems to have a negative back wash effect.

Some Final Remarks

Authorities in the area of language testing agree that oral production in communicative situations is one of the most arduous tasks to evaluate. This is because "the elements of speaking are numerous and not always easy to identify. [Consequently, it is largely pondered over] as the most challenging of all language exams to prepare, administer and score," (Madsen, 1983, p. 147). Problems related to practicality i.e. resource, time and money needed, having no well defined insight about the nature of speaking skill, eliciting appropriate speech samples and evaluating these samples and having no reliable analysis for the purpose of objective testing are some of the sundry challenges that a tester faces with regard to oral assessment performance (Hughes, 1988 and Madsen, 1983).

Our attempt to incorporate speaking evaluation in a context where it was given little emphasis was a commendable practice. Reflecting on the intervention is also crucial informant of ones own action which shapes similar endeavor. Some of the lessons I have drawn and the other colleagues of mine should, probably, draw include the following.

The impromptu speech we used as a speaking evaluation format seemed to be incongruous with the classroom speaking activities, and less reliable. It also seemed to have questionable   beneficiary wash back effect on the teaching and learning process in the context under question and seemed to snatch too much of the last precious instructional, tutorial and study hours for it was administered around the end of the semester.

The implications that could be drawn from this retrospective reflection include

- Micro skills to be assessed need to be identified prioritized, and assessment instruments which can clearly mirror the testees' abilities in these constructs should be selected and used;
- Test formats should go hand in hand with the formats practiced in the course;
- Continuous assessment seems to give a relatively more reliable picture about testees' performance than a single test like the impromptu speech and continuous assessment may enhance intra rater reliability; and
- Though practicality problem may discourage two or three instructors to go to a class to assess students speaking performance, involving students in the assessment process with guidelines, may not only increase inter rater reliability but also motivate students.

References

Alderson, C., Clapman, C and Wall, D. (1995).**Language Test Construction and Evaluation.** Cambridge: Cambridge University Press.

Atkins, J., Gebremedhin Simon and Hailemichael, Abera.(1996). **College English.** Addis Ababa: Addis Ababa University Press.

Bachman, L.F. (1990). **Fundamental Consideration in Language Testing.** Oxford: Oxford University Press.

Cummings, W. K. (2003). **The Institutions   of Education**. UK: University Press.

Hughes, A. (1989).**Testing for Language Teachers**. Cambridge: Cambridge University press.

Madsen, H. (1983). **Techniques in Testing**. Oxford: Oxford University Press.

Weir, C. (1988). **Communicative Language Testing**. UK: Prentice Hall International Ltd.