

Assessing Speaking: Issues and Concerns

Tulika Prasad

Associate Professor, Satyawati College Eve. (University of Delhi)

Email Id: tulika_40@yahoo.com

Abstract: This paper will focus on the construct of assessing speaking in a multilingual context. It will initially discuss the need to test speaking, then explore the 'what 'of speaking, i.e. what is it that we want to assess in speaking skills and finally the 'how 'of speaking i.e. the need to construct valid and reliable test items, developing valid criterion and scales to measure speaking

Keywords: Speaking, language skill, language function, proficiency, tests, tasks, rating scale.

Introduction

It is generally believed that among the four language skills, speaking, particularly in a foreign language, is the most difficult language skill to assess. The various directions and foci in the testing of speaking abilities of learners frequently lack solid grounding on theory and pedagogy and reliable test designs. (Pennington, 1999; Celce-Murcia, Brinton, & Goodwin, 1996).

1. Why test speaking?

'Speaking skills are an important part of the curriculum in language teaching, and therefore this makes them an important object of assessment as well'. (Sari Luoma: 2006). The objective of teaching spoken language is the development of the ability to interact successfully in that language and that this involves comprehension as well as production. However, there are few opportunities to evaluate students' speaking skills because speaking assessment is not administered in any formal exams including the high school final exams as well as exams at the tertiary level. . The reasons could be many ranging from lack of infrastructure, administrative and financial costs, lack of adequate trained testers, shortage of time, logistics, its absence in the

syllabus and of course large numbers. Since oral skills are neither taught or tested language teaching and learning is incomplete, rather distorted.

2. What to test in speaking?

What we want to test basically are various language functions through a variety of test items. Once we agree on the fact that it is important to assess speaking, the next question that we need to ask is what is it that we want to test in speaking and the specific aspects of speaking that we want to assess at different levels of proficiency. Moreover we also have to make sure that the tasks set to evaluate speaking should elicit behavior which truly represents the candidate's ability and which can be scored validly and reliably.

In testing the speaking proficiency of the students we are mostly interested in testing the language abilities of the students in various real life situations/contexts in their social and personal life. It is therefore imperative that our test tasks mirror the important features of real life situation, and be as realistic and direct as possible. Wherever possible we would like the test tasks to be as direct as possible, incorporating as many of the critical features of real-life language use as possible. To achieve this the first thing we need to ensure is that the conditions under which the test is to be conducted is as real as what we encounter in real life situations. We also need to contextualize the test items so that students are expected to perform in authentic situations, i.e. the tasks should be as authentic as possible in realistic contexts. Moreover this will also ensure a positive washback on the teaching that precedes the testing, since the teacher while teaching speaking skills will ensure that the students are exposed to tasks which are authentic and realistic. Apart from this when designing any test it is important to know what we want to test or which specific skill of speaking we propose to test through the task. This will not only help the assessor but also the interlocutor to help assesses to focus on that particular aspect of speaking. We also need to ensure that the particular language ability we propose to test must become explicit through the test task. If this is not so, then the task becomes redundant/task is not valid as it has failed to test what it proposed to test.

Possibly some of the language abilities we would like to assess at the basic level could be giving and asking for information, narrating incidents, describing events and pictures, participating in basic conversation around a familiar topic, describing their likes and dislikes, plans and arrangements etc. whilst at a higher level the expected language proficiency that we

would expect students to achieve could be talking about their hobbies, aspirations, dreams, achievement, asking/giving instructions, directions and participating in debates in class, engaging in conversations with peers and others on topics of interest, talking about experiences, beliefs, values coherently and confidently etc.

Once we have agreed upon the various language abilities we would like to assess the next parameter that we need to consider is the conditions under which these abilities will be assessed. It is important that a context (linguistic and sociocultural) needs to be built if exchange of language is to be meaningful. Therefore a conscious effort has to be made to build into test items real life contexts for assessing the ability in real time. To measure language proficiency adequately in each situation, we should take into account why the language is to be used, where, how and with whom, on what topics and with what effect. The setting of the task also needs to be given due considerations. If the candidates are placed in a setting, say for a role play, which they may not encounter in their future life, the task is not considered to be valid. Of course complete authenticity of setting is obviously not attainable in the classroom, but the setting should be made as realistic as possible and the roles given to the students should certainly be within their experience and appropriate to their age and culture.

3. How to test speaking?

Once the basic abilities in spoken language that we wish to test have been identified or agreed upon, the next step would be construction of test items to assess these language abilities.

Because speaking is done in real-time, learners' abilities to plan, process and produce the foreign language/L2/ second language are taxed greatly. Speaking is also the most difficult language skill to assess reliably. A person's speaking ability is usually judged during a face to face interaction in real time, between the interlocutor and the assesse. The assessor has to make instant judgments about a range of aspects of what is being said. This means that the assessment might depend not only upon particular features of speech e.g. pronunciation, fluency, accuracy, but the assessor also has to pay attention to a host of other factors such as the language level, gender, the personal characteristics of both the interlocutor and the candidate. Moreover, the nature of interaction, the sort of tasks that are presented, the questions asked, the topics broached, and the opportunities that are provided to show his/ her ability to speak in a foreign language will have an impact on the candidate's performance.

While preparing a good test it is important to keep in mind the features of a good test. Therefore while developing a valid and reliable test we need to ask ourselves the following questions:

- a. Are the tasks set at the appropriate level of difficulty?
- b. Will the tasks discriminate adequately between the performances of candidates at different levels of proficiency?
- c. Does the test assess the full range of appropriate skills and abilities, as defined by the objectives of the course?
- d. Is there an overlap in the structures, skills in the different tasks?
- e. Are the tasks and the instructions clear, so that the examinee knows exactly what is expected of them as a response?
- f. Can the tasks be completed satisfactorily in the allotted time?
- g. Are the texts and tasks culturally neutral? Moreover do the tasks have varying formats and preferably does each task make different demands from the learner?

(Adapted from 'Understanding Language Tests' Cyril Weir)

As discussed earlier all the tasks should be direct and incorporate language use in real life contexts.

Developing a valid measuring/rating scale and criteria: Issues and concerns

North (1996) describes the challenge of developing rating scales as ‘trying to describe complex phenomena in a small number of words on the basis of incomplete theory’. Furthermore, as Brindley (1998:116) notes, it is not always easy to tell what scale descriptors are meant to describe – what learners ought to be able to do at each of the scale levels or what they actually do.

Scales are difficult to write both because there isn't much evidence about language learning and because of the need to summarize descriptors into short statements to make them easy to use. It is also legitimate for scales to differ depending on their main purpose and their target audience, and in fact it is useful to make different versions for different audiences rather than making one version to suit all purposes. (Alderson, 1991: North 1996;) It is generally

believed that if we can develop scales at a number of levels, the progress recorded will be more specific. But to develop scales at many levels requires consistency on the part of the assessors which is not easy and therefore it is advisable not to have more than four to six scales.

The criteria as defined in the scales generally talks about measuring the four skills with not more than four to five descriptors per skill per level. While trying to develop appropriate scales it is a good idea to refer to an already existing scale to develop a good and appropriate scale to assess speaking skills. It is also appropriate to have positive sentences as descriptors especially if they are learning goals. The Can Do statements for language learning as described in the CEFR (Common European Framework of Reference) scales can be referred to while developing a scale. However these scales must be related to the purpose of the test and the definition of the construct to be assessed.

To develop good clear and concrete descriptors one has to have a clear concept and understanding about the desired proficiency levels of the students at the various levels of the course which can be translated into concise and accurate descriptors/ sentences. The basic rules for writing good skill level descriptors are that they should be brief, clear, definite, and comprehensible independently without reference to the other descriptors. Brevity makes scales user-friendly for the assessor and for those who are reading the scales for the first time. Simple words and sentences should be used while writing the descriptors to help the readers of the scales to understand it well and these sentences should be clear and concrete.

Concerns in developing appropriate tasks to assess speaking:

When we plan to develop a speaking task to assess the oral proficiency of students we can develop a good test by having a model which we can refer to, rather relating the test to a model of language ability. Therefore when we sit down to prepare test items to assess the proficiency of the examinee in speaking we also need to define the construct we want to assess. Once we are clear about the construct, the test item developed can be made, keeping in mind this construct.

This implies that while developing tests for assessing speaking, we also need to have a fair idea about what kind of speaking the test will focus on, how the assessment will be done and what the rating criteria will be. These test specifications contain the test developer's definition of the constructs assessed in the test, and detailed definitions of the test and rating criteria to guide

the development of comparable tasks and the delivery of fair ratings. Though there are several frameworks that talk about test specifications , the most detailed list of contents of test specifications is found in Alderson et al. (1983: 11-20, 38), according to whom the specifications should define:

- the test's purpose
- description of the examinees
- test level
- definition of construct (theoretical framework for the test)
- description of suitable language course or textbook
- number of sections/papers
- time for each section/paper
- weighting for each section/paper
- target language situation
- text-types
- text length
- language skills to be tested
- language elements to be tested
- test tasks
- test methods
- rubrics
- criteria for marking
- descriptions of typical performance at each level
- descriptions of what candidates at each level can do in the real world
- sample papers
- sample of students performance on task

While Lynch and Davidson (1994) and Alderson et al (1983) proposed that all the contents listed above should be compiled into one document: the test specifications. Bachman and Palmer (1996) suggested dividing the same content into two different documents: the design statement, which will contain the background definitions for the test, and the test blueprint which specifies the structure of the test and also contains the test task specifications., such as the purpose, the construct, setting, time, instructions, and the linguistic features involved.

The third approach could be putting down the specifications in a module form. The specifications as defined in a module form consist of three or four parts.

- a. **Construct module:** this defines the skills to be assessed. It also defines the relationship between the abstract definitions of the skill to be assessed and the concrete implementation of them in the tasks and criteria.

- b. **Task specification module:** this talks of the tasks that the assessment consists of and also specifies the information and the instructions to be given to the test takers. It also discusses in detail each of the task items and the skills which each will assess.
- c. **Assessment specification module:** this defines the rating criteria and specifies how they are to be used while assessing.

The advantage of writing the test specifications is that they help the test developers in creating a coherent system of the test construct, task and assessment criteria.

Apart from writing out the test specifications for each test, all tests need to be both valid and reliable. A test is considered reliable if the scores on the test are consistent irrespective of them being administered to the same people a number of times or on different days. Reliability of a test is important since it implies that the scores can be relied upon while taking decisions on either the test takers or the test. While developing a test, validity is considered an important parameter for a good test. A test is considered valid if it clearly defines what the test proposes to measure and consequently the test outcome shows that it has measured what it proposed to measure. For e.g. if the purpose of a test item is to measure a student's ability to narrate a story in a coherent and organized manner and if the assessor is able to measure this when the task is administered, the task is considered to be valid.

i. Examples of some test formats to assess speaking and features of the spoken language the task will assess:

a) Sentence repetition

The student is expected to repeat statements exactly what s/he hears on the tape. The statements could vary from single words to long sentences depending on the level one is assessing. Although the reliability of this kind of testing is high, it still is an indirect measure of a student's speaking ability. It is difficult to translate performance on this test into statements about an individual's language ability.

b) Giving a situation and asking students to respond to it

Example: You have lost your library ticket and need information as to how to get a new library ticket issued?

You have to go to the bank to deposit a cheque. You want to find out from the bank if they are open at 4 p.m.

This task exposes students to a variety of real life situations and a variety of speech events. Moreover all the students receive the same kind of input so this task measures high on fairness. If there are 2 teachers then both can mark the student simultaneously, and chances of bias or error is taken care of.

c) Telling a story with the help of pictures

Students are given some pictures which are in the correct sequence and expected to describe them or tell a story preferably in the past tense. This task expects students to talk continuously therefore gives an opportunity to the assessor to measure their proficiency on a range of criteria such as organization, coherence, use of past tense, use of appropriate vocabulary etc.

d) Interaction between 2 students/ Interaction in pairs

The students are given a situation and then given role cards. Depending on the level, these cards may have some pointers/statements to help/ cues to support students. The students are expected to discuss the situation and then role play it. They are expected to communicate with each other and each of them is more or less given equal time to speak. They are provided an opportunity to exchange information, ask and answer questions, etc. This task provides for a lot of language use and helps the assessor to evaluate effectively the students language ability on aspects of coherence, organization, use of appropriate vocabulary, grammar, improvisation. Real life situations/contexts are incorporated in the task. Students are expected to solve problems, arrive at conclusions, and even argue out their position on an issue. The interaction is purposeful and unpredictable to a certain extent. The candidates have to negotiate meaning, and monitor and respond to the discussion spontaneously. Use of paralanguage, gestures, pauses are all incorporated in this kind of a task. The task also demands turn taking and ways to signal when they want to speak, recognize the right moment for taking a turn, know how not to lose their turn, allow the other speaker his turn.

One drawback of this task is that if one of the speakers is dominating and does not allow the second speaker to talk it will not be possible to assess correctly the language ability of both the

speakers. This format allows the test maker to construct many such situations to be used for evaluating speaking.

e) The unstructured interview:

This is a popular format used precisely because of its face validity and content validity. The interview is like an extended conversation with no fixed direction or agenda. The candidate has an opportunity to talk of issues he/she is comfortable with and is equally responsible for continuing a meaningful dialogue. The flexibility of this task is its major strength and depending on the scope, need and level it can be modified. One of the negatives about this task is that it may not cover a range of situations which students may encounter in real life. This format doesn't work well with students who are shy or reticent. But this task is time consuming and difficult to administer if there are many students. The examiner for this kind of a format has to be very experienced because assessing is not very easy in this format.

f) The controlled interview / structured interview:

In this format the set of questions to be asked to the examinee have already been decided beforehand as well as the language ability to be measured. The examinee only speaks in response to the questions asked. The student has to process information and respond in real time which gives a fair picture of the student's language ability. The student is expected to talk about various things and cope with talking on various topics. This kind of a format allows students to respond to more or less the same set of questions and so it is more fair for the students. For the examiner it is easy to make comparisons across performances and therefore this format scores high on reliability.

g) Comparing and contrasting pictures:

Students are given 2 pictures and asked to compare and contrast them. These kinds of tasks are more demanding than simply describing a picture. They require analysis and a discussion on the similarities and differences, use of comparative forms, and some difficult grammatical structures.

h) Decision tasks/ problem solving tasks :

The students are given a problem and are required to resolve it or take decisions on how to solve it. Students are expected to express their opinions and also justify their decisions. These tasks

can also be interactive or done in pairs or in a group. They will also be expected to take turns, negotiate, convince others about their decisions etc , The content and cognitive level of this task is quite high.

i) Role plays and simulations:

Two students are involved in this task. Each of them is given a role to enact. Each of them is explained their role and then expected to simulate and carry on the act. In this kind of a task examinees take on a new role and their performance shows their ability to adapt to the requirements of the new role and situation. Of course the situation must be as close to real life as possible. This task will also reflect the students use of a variety of language resources.

5. Conclusion

Even though I have tried to build a framework which can be used to assess oral proficiency, yet assessing speaking is challenging, because there are many factors that influence our impression of how well someone can speak a language, and also because we expect test scores to be accurate , just and appropriate for our purpose. This is not an easy task, though both teachers and testers have tried to achieve all this through a range of different procedures.

References

- Alderson, J.C.(1983).Who needs jam? In Hughes, Aand Porter, D. (eds). *Current development in Language Testing*. London: Academic press . pp 87-92
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Brindley, G. (1998). Describing language development, rating scales and second language acquisition. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between SLA and Language testing research* (pp. 112-114). Cambridge, England: Cambridge University Press.
- Luoma, S. (2003). *Assessing Speaking*. Cambridge: Cambridge University.
- Lynch,B.K., & Davidson, F., (1994). *Language Test Development: Linking Curricula, Teachers, and Tests*. Tesol Quarterly, Vol 28 , No 4.
- North, B. (1996/2000). The development of a common framework scale of descriptors of language proficiency based on a theory of measurement. Ph. D thesis. Thames Valley University/New York: Peter Lang.
- Pennington, L (1999). *Assessing the communication skills of children with cerebral palsy: does speech intelligibility make a difference?*
- Weir, C. (1990). *Communicative Language Testing*. New York: Prentice Hall.